Pragmatic Causation

Antony Eagle
Oxford University & Exeter College
antony.eagle@philosophy.ox.ac.uk

In Huw Price and Richard Corry (eds.) *Causation, Physics and the Constitution of Reality: Russell's Republic Revisited*, Oxford: OUP (2007): pp. 156–90.

Abstract Russell famously argued that causation should be dispensed with. He gave two explicit arguments for this conclusion, both of which can be defused if we loosen the ties between causation and determinism. I show that we can define a concept of causation which meets Russell's conditions but does not reduce to triviality. Unfortunately, a further serious problem is implicit beneath the details of Russell's arguments, which I call the *causal exclusion problem*. Meeting this problem involves deploying a minimalist pragmatic account of the nature and function of modal language. Russell's scruples about causation can be accommodated, even as we partially legitimise the pervasive causal explanations in folk and scientific practice.

§1 Russell's explicit arguments against causation

Russell (1913) famously attacked the law of causality, and indirectly the concept of causation itself, as an inessential anachronism encumbering our proper understanding of the world disclosed to us by science. Russell's primary target is the *fundamental law of causality*, which is the principle that every actual event has an actual sufficient cause, one that is guaranteed to bring that event about and in fact did so. He proposes to argue against the law by indicating that the causal relation between events that it requires does not exist.

For Russell, the relation of causation is the relation of determination: *c* causes *e* just when *c* determines *e* to occur. This succeeds as an analysis only if the relation of determination is antecedently clear. From Russell's discussion, it is apparent that he thinks determination requires that the occurrence of *c* be sufficient for the occurrence of *e*; we shall assume, not implausibly, that *c* determines *e* just when the fact of the occurrence of *c*, in conjunction with the background laws $\mathcal{L}$, implies the fact of the occurrence of *e*.[1] Different laws give different determination relations; we assume here that the laws of interest for the purposes of analysing causation are the laws of nature.

These definitions of causation and the 'fundamental law' are not a Russellian idiosyncrasy. The idea that causation is determination of one event by another appears in Hume's 'constant conjunction' regularity analysis (if *c* and *e* are *constantly* conjoined, the occurrence of *c* should be sufficient for the occurrence of *e*). Even some later accounts of probabilistic causation, which are skeptical about the supposed 'universal law', retain the idea that where there is causality, the causes determine their actual effects (Suppes, 1970). These views must therefore either deny that every event has a sufficient cause in order that the existence of causality might remain compatible with irreducibly probabilistic theories, or else join many physicists in bemoaning the 'disappearance' of causality from indeterministic quantum physics. This close connection between determination and causation remains a unifying feature of many otherwise conflicting accounts of causation (Norton, 2003: §2)—causes *make* their effects happen in some way that doesn't

[1]This derives immediately from the slightly more general semantic condition that models of the laws contain the determining event only if they contain the determined event (Earman, 1986).

leave the effects (wholly) up to chance.

Russell thinks that causation as a relation of determination between events doesn't appear in fundamental physics—indeed, cannot be rendered compatible with fundamental physics—and hence should be jettisoned from a properly scientific world view. From our perspective, the indeterminism demanded by standard quantum theory might already seem to undermine this conception of causation. Russell's arguments, if sound, would raise problems for causation even supposing a deterministic fundamental physics.

Following Field (2003), we may identify two arguments in Russell to his conclusion. The first rests on the claim that the equations of fundamental science are *bi-deterministic*. That means the state s of some system S at t fixes the whole trajectory of S through the space of possible states both before and after t, in the sense that any two systems both in s at t would share all their states at all times (Earman, 1986). If we make the plausible supposition that each macroscopic event is constituted by some particular microphysical state, then fixing on some particular event in the system at a time will establish 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' is defined in Russell's sense—namely, c determines e—then by this argument e equally well causes c. However, causation is, intuitively, asymmetric: if c brings it about that e, it is not the case that e has any causal influence on c. In bi-deterministic physics, the asymmetry of determination, and hence the asymmetry of cause and effect, is lost. There seems no place in bi-deterministic physics for the causal asymmetry, which is a fundamental feature of the intuitive concept we are attempting to analyse.

This first argument isn't very compelling, for at least a couple of reasons. Firstly, though it might be true that the causal asymmetry is not an asymmetry of determination, causation still might be defined from a relation of determination *combined with* an asymmetrical relation, where the asymmetry comes from somewhere else. One suggestion is that the asymmetry of causation depends on the asymmetrical temporal distribution of entropy, which increases with time due to unusually low entropy initial conditions (Albert, 2000).

Secondly, it may still be the case that there is a macroscopic asymmetry be-

tween 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 *localised* event c, whose character does not determine every aspect of the global state, presumably there are many global states that can involve this type of event as a constituent part. Even in a deterministic system, it is possible that both (i) all trajectories which feature this event type c as part of some global state s at t have some further type of event e as a feature of a state s' at t', and (ii) not every global state that features e at some time lies on a trajectory which involves some past state that features c. 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 events of a purely local character, rather than the entire state of the system, might very well give us an asymmetry of determination between particular events. Moreover, it is arguable that this local determination is exactly what the original notion of a cause was supposed to capture, since the typical situations we use to elicit our causal intuitions concern kinds of medium-sized events, not time slices of the whole world. This second objection, which I find very compelling, depends on our being able to provide a satisfactory account of the concept of a local event.[2] For our purposes, it seems most important that the event in question must be of restricted spatiotemporal extent. Our commonsense causal views also seem to require that the events must be *discrete* (readily distinguishable from other events going on around them), and therefore easily subsumed within familiar and natural (to us) delineations of the categories of events. Other features may also be involved, though I am unable to do full justice to these issues here.

Russell must have been sensitive to this weaknesses of his first argument, since his second argument picks up on this conception of causation between local events. Consider some small local events c and e such that the occurrence of c determines the occurrence of e but not vice versa. Russell argues that these 'local' events won't be the kind of things we typically take to be related by cause and effect. That is, events that enter into a relation of determination will not typically both be discrete and non-gerrymandered events. But then we have saved the idea of causation as determination only at the cost of savaging our ordinary intuitions

[2]This, in turn, is tied up with characterising the physically important concept of *locality* (Lange, 2002).

about the kinds of events that can be related as cause and effect. The physical relations between states which underlie the determination relation between local events do not hold between the kinds of events that feature in our causal intuitions, or, most importantly, in our causal reasoning and the actions that flow from that reasoning.

As an example of Russell's point, consider the causal relations that support ordinary attributions of blame and responsibility. For example, suppose we blame Slim for killing Bruce by shooting him. Rather obviously, Slim's firing of a gun F and Bruce's death D are not of the right kind to get into the determination relation, because F can occur without D: for example, if Slim had missed, or if Sharon had thrown herself in harm's way, or if the bullet had exploded harmlessly in mid air. (Similarly, in cases of preemption D could occur without F.) The problem with local determination that this case discloses is that there is always the chance of some *interference* from outside the local area at the 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 set of events, perhaps the whole past cone of potential causal influence on the effect. So if c really determines e, c will have to be incredibly more complex and larger than causes are typically taken to be. Countless other factors must be accounted for so that the natural and salient cause can determine the effect in question.

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. Even if we allow 'negative' events, like the non-occurrence of a potential defeater, some positive characterisation of what is occurring at the potential locations of those defeaters must be possible, and that characterisation is important for the derivation of the determination relations from the fundamental physics. Crucially, any genuine and robust relation between cause and effect that is derived from the underlying physics will involve some events that are intuitively not causes being counted as causes simply by virtue of their being in a potentially causally efficacious location. (Precisely analogous lines of argument hold for the generation of spurious effects.)

Indeed, we shall be unable to make a distinction as regards causal efficacy with respect to these events—in particular, we shall not be able to distinguish

the genuine from the merely potential causes from a set of events each of which occupies a particular kind of location, if all we have to go on are the relations that physics gives us. All of these causes and potential defeaters are necessary for the laws of physics to determine the effect consequent upon the cause; leave any out, and the determination relation will no longer hold. But this makes no distinction between the events that were, intuitively, really the causes, and those that contributed merely by having some character or other that failed to interfere. Often this kind of problem is solved by an appeal to explanatory or contextual salience: that, strictly speaking, they are all causes, but one is singled out for explanatory purposes because of special particular relationships that we bear to it. This may well be so (indeed, I'll argue for it in §4), but we must recognise that the salience of some event appears nowhere in its purely physical description, and no appeal to the underlying physics will establish any priority for the genuine causes over the spurious causes.

The importance of this emerges when we consider causal deliberation. If, as Russell's arguments suggest, we can't distinguish a genuine cause from an actual non-cause that might have been a preventer, then we shall be unable to engage in goal directed activities that depend on effectively bringing about certain states of affairs. This is partly because the set of events we shall have to manage or intervene on is far too large and inhomogeneous to deal with effectively. The real problem, however, is that we shall be misled into performing actions to bring about our goals that are not effective for achieving those goals. Or rather, physics provides no reason why altering the situation with regard to genuine causes should make any more sense than altering the spurious causes. Of course, we have intuitions that govern our causal interventions, and those intuitions serve us well. But they get no support from fundamental physics, if Russell is right. Cartwright (1979) emphasised this aspect of causation when she talked of effective versus ineffective strategies. Russell can be seen as turning this proposed justification for causal reasoning against itself: if we really want causes to determine their effects, then intuitive causal reasoning simply isn't adequate for distinguishing effective versus ineffective strategies. We need the full power of our best physics; and unfortunately, our best physics doesn't provide a determination relation that meshes at all well with our folk theory of causation (Norton, 2003: §5).

We can put the net result of these arguments as follows. 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 we wish to apply the concept of causation to some spatiotemporally local situation, then we can only have determination if we are willing to abandon the role of causes in demarcating effective strategies for manipulating that situation, which is to say if we abandon the traditional concept of causation altogether. The genuine physical determination relation that really meets the demand for effective strategies renders any use of an additional, fallible and intuitive conception of causation redundant. If the notion is redundant in physics, it is dispensable from a properly scientific account of the world. Since fundamental science is the arbiter of genuine ontology, this relation of causation should be excised from our folk ontologies: it can't even be reduced to physics, let alone found within it.

It seems to me that this eliminativism is precisely the wrong conclusion to draw. The real fault lies in the conception of causation as requiring determination. When this is combined with Russell's observations concerning the kinds of determination relations that can be defined within fundamental physics, it follows that the relation of causation is trivialised or rendered unrecognisable, forcing us to include merely pseudo-causes. The appearance of triviality doesn't show that the folk conception of causation is to be rejected; rather, it shows that one part of that folk concept—determinism—is in tension with the rest of the notion. If we could show that we could give an analysis of this folk notion of causation that limited or restricted the demand for determination of effects by causes, we could respond effectively to Russell's eliminativism. We wouldn't have a 'universal law' of causality, that every effect has a determining cause. But that cannot be had at all.

In the following section (§2) I sketch one such account of causation that weakens the determination requirement and yet is adequate to the other tasks of a concept of causation, such as grounding the distinction between effective and ineffective strategies. Before going on, I will briefly sketch the plan for the rest of this essay. In §3 we will see that a particular argument, which I call the *causal exclusion problem*, lies beneath the details of Russell's discussion, and that argument threatens to render the reconstructed concept of causation otiose. In responding

to this fundamental objection in §4, I shall appeal to broadly pragmatic considerations grounding the legitimacy of certain linguistic practices—the facts supporting the legitimacy of causal language are outlined in §4.2.

§2 Causation as partial dependence or determination

In this section I give a sketch of my favoured analysis of causation without determination.[3] Though I think very highly of the account, my main concern here is not to defend it, but rather simply to demonstrate the possibility of providing an adequate causal relation that evades Russell's arguments. That the account sketched here can do so should count in its favour against other accounts.[4] Be that as it may, the purpose of the following is illustrative rather than an attempt to argue for the framework outlined here. Those who dispute the details of this approach should feel free to substitute into my argument at this point their own preferred account of causation, provided they can satisfy themselves it too will avoid Russell's arguments. However, it will turn out that some features of this account, such as its emphasis on manipulation and intervention, fit beautifully with the eventual pragmatic justification for causation in §4.2.

§2.1 COUNTERFACTUALS AND PARTIAL DETERMINATION

This account begins by proposing that we analyse non-determining causal dependence between local events in terms of relations of counterfactual dependence be-

[3]The broad outlines of the technical aspects of this approach should be familiar from the work of Pearl (2000) and Spirtes *et al.* (2000). The philosophical development of a counterfactual theory in the causal modelling framework is basically similar to Hitchcock (2001) and Woodward (2001), though in this context I stress, for obvious reasons, the role of counterfactual dependence as being essential to avoiding the problems raised for full determination by Russell's arguments. For more philosophical detail I recommend Woodward's very rich recent book (2003). There may be an interesting connection also 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 account may capture this, as well as yielding counterfactuals which give the influence a uniform treatment within the counterfactual framework.

[4]For instance, I suspect that the conserved quantities theory of Dowe and Salmon will have difficulties in meeting Russell's challenge (Dowe, 2000; Hitchcock, 1995).

tween propositions concerning those events.[5] The starting point is Lewis' (1973a; 1979b) analysis of causal claims in terms of certain counterfactual conditionals. Lewis proposed, roughly, that c causes e just in case the counterfactual claim $\neg c \mathrel{\Box\!\!\rightarrow} \neg e$ ('if it hadn't been the case that c, then it wouldn't have been the case that e') is true.[6] This analysis captures the intuitive idea that causation is a species of *difference-making*: causes make the difference as to whether their effects occur or not, such that their absence would entail the absence of the effect, other things being equal. These claims about difference-makers are most plausibly rendered as counterfactual claims, and any satisfactory account of causation must make sense of the intimate connection between facts about causation and the counterfactuals made true by those facts (regardless of whether that account takes this connection to exhaust the facts about causation). We may then legitimately use these counterfactuals to explain various features of causation; in our case, we use counterfactuals to explain how causes might at least partially determine their effects.

In passing, let me note the need to impose the further constraint that $\neg c \mathrel{\Box\!\!\rightarrow} \neg e$ is not a *backtracking* counterfactual. Paradigm backtracking counterfactuals are those that are true under an *evidential* reading of the counterfactual connective as reasoning from symptoms to cause, or from evidence to hypothesis. For instance, 'if it had not been the case that the glass was broken, then it wouldn't have been the case that I smashed it earlier' is true, taking the unbroken glass as evidence for the claim that the glass wasn't smashed. Of course, though this conditional fits the pattern $\neg c \mathrel{\Box\!\!\rightarrow} \neg e$, we cannot take it to express a true causal claim. These conditionals are called backtrackers because characteristically the consequent temporally precedes the antecedent. Insofar as this temporal asymmetry is an intuitive feature of causation, so excluding backtrackers helps capture our platitudes about the relation between causation and time. Having indicated the potential danger of backtracking counterfactuals, I will henceforth assume that all the counterfactuals discussed below are normal.

Earlier, we assumed that the determination relation involves the occurrence of

[5]It is especially nice because it preserves a kind of defeasible determination, in a way that, for example, probabilistic analyses of causation do not.

[6]Here of course I've used the same notation for the event c and the proposition which says that c occurs, the latter appearing in the counterfactual conditional. I trust no problematic confusion will occur.

the determined event being implied by the occurrence of the determining event, given the laws. Ordinary conditionals, such as those characterising the implication relation, satisfy the following inference rule: $\alpha \rightarrow \beta \vdash (\gamma \wedge \alpha) \rightarrow \beta$ (the rule of 'strengthening the antecedent'). If the conditional appearing in statements of determination is a standard conditional, as we assumed, then the relation of determination holds regardless of what else happens to be true. That means α must fix every fact relevant to the occurrence of β in order for the determination relation to hold: the only facts it need not fix are those irrelevant to the holding of the relation, like γ in the inference rule. This feature of the logic of the determination relation gives rise to Russell's problem of spurious causes: every relevant potential cause must already be fixed or entailed by α, and hence counted as part of the cause.

Counterfactual determination might appear to share this feature with standard conditional analyses of determination, especially if we attend to the gloss "if it were the case that *c*, then it *would be* the case that *e*". But that appearance is misleading. Counterfactuals do not support strengthening the antecedent, as the following series of counterfactual claims shows. Imagine that Slim shoots Bruce. Were Slim not to have fired his gun, Bruce wouldn't have died. But were Slim not to have fired his gun *and* had Sharon fired her gun, Bruce would have died. But were Slim not to have fired his gun, and Sharon had, *and* Sharon had missed, Bruce wouldn't have died. And so on. All of these counterfactual claims are true, both intuitively and on the standard Lewis-Stalnaker semantics. We can see therefore that counterfactual dependence is *not* invariant if we place additional conditions in the antecedent. These additional considerations are typically potential events that we didn't consider in the initial attribution of causal effectiveness to the antecedent event. If we analyse causation as counterfactual dependence, then we may have true statements of partial determination of effect by cause while also leaving open the possibility that there may be a recherché situation where the cause is compatible with the non-occurrence of the effect. This kind of determination relation, I submit, is the best we can have that manages to avoid the Russellian spurious causes argument, retain the connection between causation and determination, and preserve the intuitions surrounding determination. Causation, on this view, involves partial or defeasible determination, as captured by counterfactual

dependence.

One situation where full determination does occur even on this analysis is if the antecedent and consequent events are *maximal* global states of a system, such that either φ or $\neg\varphi$ is fixed by the state Γ for any φ. Then $\Gamma_1 \boxdot\!\!\rightarrow \Gamma_2$ cannot be defeated by affixing additional events, for there aren't any. Russell's first argument can be restated: given the fundamental physics we have, for any two global states such that $\Gamma_1 \boxdot\!\!\rightarrow \Gamma_2$, there is a true counterfactual $\Gamma_2 \boxdot\!\!\rightarrow \Gamma_1$. There is thus no asymmetry of determination, and hence no causal asymmetry. Russell's second argument is that global maximal states are too large and inhomogeneous to be the relata of the intuitive causal relation we were trying to analyse.

§2.2 A MODEL OF CAUSATION

I begin the task of integrating these facts about counterfactuals into our model of causation by making some trivial observations. There are some events that make no difference when added to the antecedent of a true causal counterfactual. Example: we don't think that were Slim not to have fired, and were some small event φ on Pluto to have occurred, then Bruce would have died. This is true regardless of what φ is. Some events, no matter how they might have turned out, are not capable of affecting the counterfactual dependence between two other events because they are isolated from or irrelevant to those events. On the other hand, some possible events ψ on earth might well have altered the truth of the counterfactual if added to the antecedent conditions. For instance, some action ψ of Sharon's in the near vicinity of Slim's firing and Bruce's being shot might have turned out in such a way as to contribute to or detract from Slim's action's bringing about Bruce's death, whether or not Sharon actually performed one of the particular subclass of actions that would actually have had an impact.

To account for these platitudes, I think an analysis of causation must have three features: first, it must relate variables (in some way) rather than arbitrary events; second, it must include a variable as relevant if some of its values are counterfactually relevant; third, it must use contextual cues to rule out irrelevant variables. Let me expand on these features.

First, the causal relata. Normally, events are taken to be the causes and effects.

The identity conditions for events are, plausibly, very fine-grained: two events are distinct if *any* aspect of them is different. So Bruce's death is different if he is shot by Slim *and* Sharon rather than by Slim alone. I am quite happy with finely individuated events. But since I believe Slim would have played a part in causing *something* (something involving Bruce's death) regardless of Sharon's action, I had better take the relata of the causal relation to be something other than arbitrary fine-grained events. I opt for a special subset of events: those involving some *random variable* taking some value. A random variable is a function from possible events to numbers, where the numbers characterise certain features of the event. For example, I could characterise the relevant class of events that might have resulted from Slim's act, and use the random variable **Death** which takes value 1 if Bruce died, and 0 otherwise (and is undefined on events that aren't relevant—where relevant will be spelled out below). This concept 'smooths out' variations between events which do not give rise to a different value of the random variable; however, some other random variable might be sensitive to some of those variations.[7] So in this case, the event of **Death** = 1 is caused by the event **Slim** = 1. The causal relata are not variables themselves—they are functions, not physical entities—but having variables involved means that it will not normally be the case that arbitrary disjunctive or otherwise gerrymandered events will be causally active.[8]

Second, what matters to the causal importance of a 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 Sharon's possible actions form a relevant class of events in our example, because some of those events, if specified, alter the counterfactual dependence between **Death** and Slim's action. 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 (while $\neg F \mathrel{\Box\!\!\rightarrow} \neg D$ also remains true). But for Sharon's activity to be *potentially* causally relevant, it doesn't matter that *actually* it was not, and she stood idly by. We might say

[7]This is what Field (2003) calls a 'fairly inexact variable', and is essentially the same notion of variable that appears in Hitchcock (2001); Pearl (2000), and Spirtes *et al.* (2000).

[8]So, for instance, we don't run immediately into puzzles that plague simple counterfactual theories which claim that some 'negative events' (those that occur just when some ordinary event fails to occur) have to be causally active, just because there is counterfactual dependence on a proposition stating that some positive event failed to occur.

that in such a case the variable **Sharon** *causally contributes* to the variable **Death**, even though the actual causal relations between actual events are insensitive to this causal relationship.

Third, we must notice that the second observation puts a necessary condition on causally relevant variables. But this cannot be sufficient, otherwise all sorts of irrelevant variables will get included because in very distant worlds they alter assessed counterfactual dependence. I think that ruling out these variables is largely a matter of contextual salience, where that depends both on background information about which variables to include (only those that have an influence in situations that are serious possibilities for us (Levi, 1980)), and also judgements about the distribution of values for those variables (Menzies, this volume)—see also §4. For current purposes I think that facts about correlation and spatiotemporal connection between events that are phenomenologically salient will go a long way to explaining the choice of variables to include in a model. On this proposal, causal relations between events are very common, but frequently it is pragmatically inappropriate to state the existence of a causal relation. With respect to the values of the variables, the role of *contrast* is particularly noteworthy: judging that the hammering smashed rather than pulverized the walnut involves making quite fine distinctions between values of the variable **HammerForce**; judging that the hammer smashed rather than failed to smash the walnut involves a coarser partition of the hammer-involving events. It is important to note that even if the variables and their ranges are chosen for pragmatic and context-sensitive reasons, the truth of the resulting counterfactuals will be a perfectly *objective* feature of those variables.

Let me deploy some new terminology to summarise the three morals about counterfactuals that I recently drew. Let the situation under consideration determine a contextually salient theory, specifying 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 facts of the form 'the value of variable $\mathbf{V}_i$ depends on the values of variables $\mathbf{V}_j \ldots \mathbf{V}_k$.' If we have a quantitative structure, the variables will be linked by functional equations that show upon which other variables the

value of each variable depends. If $\mathbf{V}_i$ neither depends on $\mathbf{V}_j$ nor $\mathbf{V}_j$ depends on $\mathbf{V}_i$, these two variables 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: (i) $\mathbf{V} \neq \mathbf{U}$; and (ii) there exists some assignment of fixed values to variables in the model such that the following counterfactual is true:[9]

Parent Were **V** to have some different value v, then **U** would have some different value u. ($\mathbf{V} = v \mathrel{\square\!\!\rightarrow} \mathbf{U} = u$.)

Note that 'grandparent' variables are not parents: if **V** only acts on **U** through **W**, then the fact that **W** is fixed on some value will prevent the change in **V** from percolating through to **U**. We say that a variable **V** is *directly causally relevant* to **U** if **V** is a parent of **U** (Woodward, 2001). This concept allows us to construct *causal graphs* as follows. Take all the variables in $\mathcal{V}$, and put them at nodes of a graph. For each $\mathbf{V}_i \in \mathcal{V}$, let $\mathcal{P}(\mathbf{V}_i)$ represent its parents. For each variable $\mathbf{V}_i$, draw an arrow from each $\mathbf{V}_j \in \mathcal{P}(\mathbf{V}_i)$ to the node $\mathbf{V}_i$. We will end up with a graph something like Figure 1. This is a qualitative causal structure, and the parenthood relations are the most basic kind of counterfactual that should be considered in causal reasoning.[10]

Consider as a simple example the model of plant growth depicted in Figure 1. In this model, **Season** is binary: growing season or non-growing season, which in turn influences the amount of sunlight (**Sun**) and the rainfall (**Moisture**). But moisture is influenced by sunlight (causing evaporation) and whether the crop was irrigated or not (**Irrigation**).[11] Finally, whether **pesticide** is used or not also influences the final plant growth (**Growth**). Some of these variables are binary, and some are quantities (sunlight and rainfall). Different models may choose to represent irrigation by a volume of water, not by a binary variable. Similarly,

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

[10]If we have in addition a probability distribution over the values of the exogenous random variables (i.e. those with no arrow leading into them), and equations which express the numerical dependence of the values of a variable on the parameters and values of its parents, we can turn this qualitative causal structure into a quantitative causal model.

[11]Irrigation and rainfall may themselves be correlated, so this model may not be perfect (since high rainfall tends to reduce the need for irrigation).

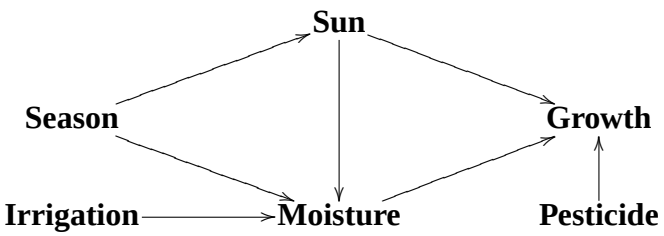

Figure 1: A sample causal model of plant growth and some of its factors.

growth may be modelled by a variable taking values of 'increased', 'decreased' or 'same'.

Interestingly, some variables can cause by 'omission' (i.e. not applying pesticide can negatively influence plant growth), and it may be noted that this framework gives an easy way for causes by omissions to work—at least, after the general problem of variable selection has been solved there is no special problem of causation by omission.[12] Of course, other irrelevant variables are also omissions of things that might have impacted, and it is the choice of variables to model along with judgements about salience and relevance that give content to our causal judgements.

Several things are noteworthy about this approach to a counterfactual analysis of causation. Consider Figure 1 again. The values of the variables **Season** and **Irrigation** are not counterfactually dependent on anything (note that the restriction against backtracking counterfactuals is crucial here: if it were not, plausibly if the value of **Sun** were different, then the value of **Season** would have had to have been different too). But presumably these variables do depend on something. To pretend we have isolated them, we can appeal to the contextual salience of local causes. Causal explanation has to stop somewhere, and if a certain condition on the parentless variables is satisfied, we should be prepared to stop with them. The simple condition is that parentless ('exogenous') nodes should be independent, not 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.) If **X** is counterfactually dependent on **Y**, and **Y** on **X**, neither through a backtracking counterfactual, then we should try to find another variable **Z** that is

[12]Thanks to Brett Sherman and Karen Bennett for help with this.

parent to both **X** and **Y** and screens off the counterfactual dependence in order to complete our causal model. (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 (Butterfield, 1992). 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 (Hitchcock, 1996).

Thirdly, we must consider what kind of facts make the Parent counterfactual true. As it stands, 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 dependent variable independent. 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.[13] The viability of the concept of an intervention depends on the possibility of *modular* causal systems (Hausman and Woodward, 1999). These are systems where each variable in the system has some independent exogenous sufficient cause.[14]

Our understanding of interventions relies on our ability to deal with quite distant counterfactual situations. Though we deploy a particular theory in reasoning about such systems, we need to be able to isolate particular parts of those systems

[13]This isn't quite right, because expanding some graph $\mathcal{D}$ 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).

[14]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.

from other parts, even to the extent of isolating the variables intervened upon from their actual causes. This will involve consideration of a set of models where the best system of laws might be quite different than the one governing the theory as a whole. How then can we get a grip on the concept of intervention? The concept of *manipulation* is important in this connection. We can begin to understand intervention by considering how human agents manipulate objects and situations in the most prosaic of circumstances. We have goals and desires, and the world is not necessarily cooperative in satisfying those desires or bringing those goals to fruition. But we have the capacity, in many cases, of doing things, of making things happen, to bring about some change in the world that makes it more amenable to our wishes. In other words, we wish sometimes to control how the world goes, and to predict what our interference might do. Once we have this simple and familiar concept of manipulation, we can begin to theorise about situations outside of our direct control, indeed those outside any human action at all. But an understanding of the way in which manipulations sever the variable intervened upon from its parents sets us up to understand how intervention more generally works, and to see manipulation as a special, particularly important, case of intervention.[15]

One problem often thought to strike accounts of causation that use the notion of intervention is circularity, because intervention itself seems a causal notion. The circularity, however, is not vicious: since the variables characterising the act of intervention itself are never included in the causal model, we may treat them as characterised solely by the counterfactuals they support. If we expand the model to include the process of intervention itself, and give a naturalised account of that particular intervention, it will no longer support the parent counterfactual, and will appear like any other causal variable. If we take the content of the concept of causation to be given by these causal graphs, the notion of intervention we need to support the Parenthood counterfactual is not a causal one at all. However, this very strategy for defusing the charge of circularity gives rise to serious difficulties in making a global causal model that includes every variable. In such a model, nothing corresponds to an intervention conceived of non-naturalistically—once

[15]See §4.2, and also Woodward (2003: esp. §3.1).

every event has a representative variable in the model, every dependence relation between events is already represented. As Pearl (2000: 350) remarks, 'If you wish to include the entire universe in the model, causality disappears because interventions disappear—the manipulator and the manipulated lose their distinction'.[16] Indeed, as I shall argue next (§3), Russell's arguments can be taken to be making a very similar point: once the global situation is considered, and every event encompassed by the model, the local concept of causation disappears.

Causation on the model described here is a relation between variables, and will apply to particular events only in a derivative fashion. This is because we have identified causation with a feature of abstract models of phenomena that can apply to many particular situations. Once we see the actual values of the variables as they appear by a particular situation, 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 (this volume), and by using 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 for some particular case. This will be a model that only applies to the single case in question. But which events are modelled depends on how we *coarse-grain* the space of events to give distinct values for the variables. For instance, the variable **Height** could have as possible values 'less than 1m' and 'greater than 1m'; or it could be a continuous distribution over some subset of the reals, say $[0, 2]$. Depending on how finely we divide up the events of 'having a certain height h', we get different causal models. This coarse-graining of the possible events is certainly a contextual factor in model construction, but it also plays a significant role in rendering causation a local relation. For example, in classical statistical mechanics the most fine-grained division of the space of possible events will make each distinct global state a different event, and we will be unable to avoid Russell's trivialising conclusions. Hence coarse-graining of events, perhaps to correspond to our epistemic and practical limitations, is necessary (in this theory and many others) for our deployment of the concept of causation.

[16]Pearl also seems to think (Pearl, 2000: §4.1) that interventions are to be connected with human free will and the causal 'unconstrainedness' of human volition.

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 causal relevance can be at times stimulatory, at times inhibitory, and so on.[17] This seems to nicely feed into Hitchcock's (2003) 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. Consider that some of the arrows might be purely inhibitory, some purely contributory, some a mix of both, and what constitutes the ground of the counterfactual claim might be very different in each case. Why think they can all be shoehorned into one neat causal metaphor like the 'cement of the universe'?

§3 The causal exclusion problem

At this point, I hope to have demonstrated the possibility of giving an analysis of causation that abandons the requirement that causes necessitate or determine their effects, without wholly abandoning all the intuitions about causation that led to the perceived connection between causation and determination in the first place. In particular, the analysis preserves the idea that causes are antecedent to effects and can be represented as such in true conditional claims; but we use a counterfactual conditional rather than the strict conditional of traditional analyses of determination.

However, a serious worry remains. Indeed, I think a version of this worry underlies both of the explicit arguments that Russell gives, and hence can be identified as Russell's real challenge to causation. The problem that threatens our account is that, even if we can show how to systematically interpret causal talk so it is *compatible* with fundamental physics, it remains true that causal talk is essentially *dispensable* once we possess fundamental physics. I call this the *causal exclusion problem*, because I think it in many ways analogous to the exclusion problem that plagues non-reductive physicalism in the philosophy of mind.[18]

[17]One consequence might be that the distinction between a cause and an enabling condition will be harder to make in this framework. Thanks to Mark Schroeder on this point.

[18]For more on the exclusion problem in the philosophy of mind, see Bennett (2003) and the references therein.

To formulate the causal exclusion problem, we shall have to tease out a couple of presuppositions of Russell's arguments. I do not think they are particularly controversial, so correspondingly I believe the threat the problem poses is extremely serious, since its presuppositions are so meagre. The first is that fundamental physics provides the justification for any true physical claim. This is not the claim that fundamental physics is complete; I don't wish to presuppose the more controversial thesis of *physicalism*, that every fact is a physical fact. Here I only assume that, if we have an account of how some physical event was brought about, or produced, in terms of other physical events, the ultimate justification for that account lies in fundamental physics. Similarly, in predicting some physical event, the ultimate guide is fundamental physics. We ourselves made this presupposition when we took care to indicate how the causal models we constructed in the last section were restrictions of more complete models of the phenomena. The counterfactual claims that I argued were fundamental to causation are supposed to be supported by consideration of how the physical situation would go, were the world in the state described by the antecedent of the counterfactual. Evaluating these counterfactuals, we tacitly presupposed, involves examining how the fundamental physics tells us how the world would go, how our best physics describes the behaviour of even quite distant possible worlds.

The second presupposition is that a complete physical account of a physical event trumps a partial or approximate account. We may well have a serviceable default technique for reasoning about various situations, justified because it leads to no significant errors for all practical purposes, or because in a limited range of circumstances it gives precisely the correct answers even though outside that range it can be in error. But in making these claims about approximations, we appeal to the fundamental physics that describes perfectly what the default theory merely approximates. The approximation gets its value from being serviceable when compared with the fundamental theory, and only has value as long as the fundamental theory continues to be accepted. If we decide that the errors it introduces are too great, or our computational capacities increase to the point where the savings gained by using the approximation are negligible, we shall have no compunction in abandoning the approximation in favour of the fundamental theory. Unfettered by practical constraints we should have no need of approximations;

and in doing ontology, particularly, such practical considerations should have no bearing on what we take to genuinely exist.

Given these presuppositions, we can state the problem. Consider some causal story in which we describe how an event *e* came to occur by citing one of its causes *c*. Is this story making a essential contribution to our understanding of the physical world, in the sense that without it, we should lack a proper account of this event *e*? If we answer that this story is making an essential contribution, then we deny that fundamental physics is authoritative with respect to predicting and describing the production of physical events. Moreover, we are simply denying the adequacy of physics, because we have a putative case in which one physical event brings about another and yet no physical explanation in terms of fundamental physics is forthcoming.

If we are not to deny the plausible thesis that physics is the final arbiter of physical questions, we must deny that this causal account makes an essential contribution to the correct description of the world. The alternatives are that it makes no contribution, or that it makes an inessential contribution. If the former, we are certainly better served by disposing of a putatively scientific concept that plays no useful scientific role. Let us suppose, then, that causal descriptions are inessential but legitimate contributions to scientific understanding. By our second presupposition, we would be better served by giving the more fundamental physical account which justifies and grounds the causal account, if our concern is with truth. The causal account is both inessential and less successful than the rival account given by fundamental physics. The relation of causation does no distinctive work in the giving of successful and accurate descriptions of the way the world goes. We should only accept the relation of causation if it can be rendered unproblematic: that is, if we can show that causal descriptions are able to be fitted seamlessly into a scientific picture of the world.

Russell's arguments enter at this point. I think they show that the concept of causation cannot be reduced to the fundamental physics in a seamless fashion.[19] The best way to reduce causation to physics would be if we could show that causation could be defined in terms of the fundamental physical relations,

[19]If I am right about this, then Norton's (2003: §4) claim that causal concepts are able to be generated by reduction to fundamental physics is not without difficulties.

those adequate for every purely physical description and prediction of the relevant events. The upshot of Russell's arguments is that, in deterministic physics, these most fundamental physical relations are relations of determination between global states, given the physical laws. If we define causation in terms of these relations, we succeed only in trivialising the concept, or in giving a reduction that clashes spectacularly with our intuitions about causation. Neither is a satisfactory situation.

If reduction fails, as Russell suggests, then we will be forced to accept causation as an additional, autonomous concept that plays no distinctive role in a purely physical description. Combine this with the notorious difficulty in defining or identifying the precise characteristics of the causal relation, and we might well think that we are better off without it. Causal descriptions simply obscure from view the real account of how any physical event happened. This obstruction of the truth is a reason why we should remove causal concepts from our repertoire; no powerful reason exists to retain of causal concepts. Hence, Russell adopts an eliminativist position with respect to causation: causation is superfluous, *excluded* from genuine physical significance. Even if they were to exist, a causal relation would at best superfluously overdetermine which events take place; without any alternative support, the fact that they would introduce implausibly widespread overdetermination should be enough to tip the scales against causal claims.

The argument is powerful. Firstly, because it works even if causation is compatible with fundamental physics. This would render our work in §2 quite beside the point, since all the elaborate counterfactual machinery we introduce only serves as a second-class kind of description. Secondly, because the assumptions of the argument are so mild, denying any of them looks at least as bad as eliminating causal descriptions, since we can replace the causal descriptions with better though more complicated physical descriptions. Denying the first presupposition, on the other hand, and introducing extra-physical causal descriptions of physical phenomena, induces skepticism about causes. Thirdly, the features of causation we have emphasised seem only to support the exclusion argument, not rebut it. Causation is context-dependent: it is sensitive to which events or variables are included in the model, and some think it is relative to default values for the variables also. Causation is partial and local. These are precisely the features which causal

accounts do not share with authoritative physical accounts, and it is precisely these deficiencies which the exclusion argument exploits. One way of putting it might be: physics has shown us how best to give an account of one event's determining another, providing a precise and detailed story where causal accounts are mysterious and imperfect. The emphasis on determination in the fundamental law of causality ironically led to the abandoning of causation as inadequate to account for determination!

Despite its many virtues, I believe the exclusion argument harbours a subtle flaw. Though causal accounts may be rendered inessential by fundamental physics, they are not thereby robbed of their distinctive contribution to *understanding*. Rather than focusing on causal descriptions of the phenomena, we should attend to causal *explanations*, which seem to play a quite different role than purely descriptive accounts. While it may well be true that the physical description of the situation is optimal with respect to the desiderata governing a good description of *how* some event came about, that same physical description needn't provide scientific understanding of *why* that event came about. Even if the global dependence relations between physical states fix entirely the physical structures, there remains an open question as to which aspects of those structures explain the resulting events. We already have *prima facie* reason to suspect that relations of global determination aren't adequate for explanation; simply observe that the these fundamental physical relations may be bi-deterministic, and so may determine the past state given the future. Yet we should rarely, if ever, wish to accept that explanation of the past in terms of the future is acceptable.

If a causal description plays any role in explanation, as seems likely, then we shall have a strong reason to think that causal explanations are not excluded by physical descriptions.[20] Indeed, it may even be the case that no account of what explanation consists in can be given independently of the concept of causation; in which case, there could not be a problem of physical explanation excluding causal explanation.[21]

If proper attention is paid to the role of causal explanations in our conceptual economy, particularly their use in hypothetical reasoning, their importance be-

[20]See, for example, Woodward (2003: ch. 5) and Lewis (1986b).

[21]Thanks to an anonymous referee for help with this discussion.

comes undeniable. Causal explanations can be vindicated by fundamental physics, in many cases, as the exclusion argument presupposes. If we can show moreover that these explanations provide a pragmatically essential handle on the physical facts, showing fundamental science to be continuous with our intuitions in some important sense, and supporting our self-conception as agents and our conception of the world as one among many possibilities, then we have provided an excellent reason to persevere with causation, despite the problems we have in giving an analysis of it. Russell's scruples about causation can be accommodated, even as we legitimise the pervasive use of causation in folk and scientific language. Or so I shall argue below.

§4 Pragmatism and the indispensability of causation

The fundamental premise needed for Russell's eliminativism is that causal accounts are never better than accounts in terms of the functional dependence between global states given by fundamental science. This premise has some initial plausibility, because of the primary role of fundamental science as providing the supervenience base for all other physical facts. But I shall argue that it is false, particularly with respect to the provision of explanations.

Let us begin by noting one premise Russell's arguments *do not* appeal to: the claim that causal explanations are incompatible with the explanations provided by fundamental physics. If this claim were true, I think it would provide an excellent reason to eliminate causal talk. But we have not relied on any such claim here (rather, presupposing it false). Nor do I think it possible to establish any such claim, since to do so would involve showing that the addition of causal claims to the body of fundamental physical truths would allow the derivation of some fundamental physical falsehood. It is clear that, on the model of causation sketched in §2, no such contradiction would be forthcoming, since each true causal claim rests on a true counterfactual claim, which in turn is evaluated by making reference to the fundamental physical models, or suitable approximations thereof. Causal models are abstractions of fundamental models; they can introduce no new claims concerning fundamental properties and relations.[22]

[22]Adding causal language, then, amounts to what Field (1980) calls a *conservative extension*

Russell's eliminativism turns more on broadly philosophical considerations rather than logical ones. The essence of his position is that causal claims are, ontologically, freely spinning cogs in the mechanism of explanations. While causal explanations may depend on physical explanations, nothing in turn essentially depends on them. A potentially dispensable element of a scientific theory which does no genuine work should, by Occam's razor, be eliminated, especially if it is a metaphysically puzzling entity in the first place.

I will quarrel with both parts of this last claim. I will argue that causal explanations do indispensable work in the social practices of giving and requesting explanations as engaged in by creatures like ourselves. I will also argue that causation stands or falls with other modal notions, so that Russell's argument actually would succeed in undermining a huge part of our conceptual framework. Without causation and cognate notions, the world would be an alien and incomprehensible place; with it, our attempts to understand and engage with our world are useful and effective.

§4.1 MINIMALISM, LEGITIMACY AND ONTOLOGY

Before we do this, however, we need to establish whether pragmatic claims about the role of some concept can do the metaphysical work we require of them.

If, as in the present case, we are skeptical about whether any fundamental facts neatly correspond to certain of our linguistic and cognitive practices, then we have two options. Firstly, we can reject the language, arguing that the only legitimate basis for it must involve it representing fundamental reality in some way. Secondly, we may reject the demand for that the language used in the practice be interpreted representationally, perhaps maintaining that the use of that language in a given practice might be understood in a perfectly naturalistic way without imagining that to every contentful sentence of a good practice there must correspond some part of reality that satisfies that content. Given these options, the causal exclusion argument I sketched above gives excellent reason for thinking that we cannot simultaneously maintain both of the following claims:

of the language of fundamental science, since no novel truths expressible solely in the original language become derivable. Of course, Field used a similar claim about number-language to argue *for* eliminativism about numbers.

Representation We must understand causal talk representationally;

Ineliminability We can maintain that causal talk is, in principle, ineliminable.

But that argument does not indicate which of these claims we should abandon.

If we had the further thought that the only way to justify a practice was to show that it correctly represented Reality, then I suppose that if we wished to maintain causal language we should be forced to posit some additional ingredient of reality, not discoverable by physics (though perhaps discoverable by philosophy?), which corresponds to the content of acceptable causal claims. Some philosophers may go where this argument leads, and accept the additional ingredients; others may deny that the practice is justified. Both call for fairly radical revisions in either everyday practice or fundamental ontology, on the basis of what must seem fairly slender evidence. Better, then, to reject the further thought: a practice may be legitimately engaged in even without corresponding to some underlying reality.

To reject the further thought is to commit ourselves to what has been called *minimalism* (Johnston, 1992: §V): the idea that sometimes a practice may be justified without that practice tracking fundamental metaphysical distinctions. Some versions of minimalism (including Johnston's) involve taking the ontology of a minimally justified practice as prima facie representational; the minimalism I will defend here will not involve this further, extremely controversial, step. Rather, the minimalism I shall defend will argue from the justification of the practice on pragmatic grounds to the claim that the language involved in the practice has a non-representational role, but that nevertheless we are entitled to use it and to act on its representational consequences. In the present section I will sketch this version of minimalism in general terms; in the next (§4.2), I will turn to causation, and argue that causal talk does have this minimal pragmatic justification. The upshot will be that causal talk is ineliminable and acceptable, yet isn't some extra non-physical ingredient of reality.

It is obvious that any adequate naturalistic explanation of the existence and persistence of a certain linguistic practice must draw upon the cognitive features of the creatures who engage in that practice. Such an explanation might involve various low level features, such as the the perceptual capacities of those creatures; or it may involve features involving higher cognitive processing, for instance capac-

ities for representation or goal-directed deliberation. Where we are the creatures concerned, it is plausible to think that there will be a complex of such features (involving our perceptual abilities, epistemic capacities, and characteristically human goals and attitudes) that will stably feature in explanations of many of our linguistic practices. Following Price (2004), we might call this stable complex of features a *perspective*, because it can be taken to characterise the viewpoint on reality that our nature forces on us (see also Price 1992; 2001). (Familiar visual perspectives originate from the constraint that we are *located* creatures.) From a given perspective, the world is most naturally modelled in ways that may not represent particularly closely the way the world fundamentally is. For example, creatures with very limited perceptual apparatus will have an internal model that will be insensitive to many features of the environment they actually inhabit.

More important for our purposes is the fact that sometimes the characteristic perspective of some creatures may make it natural to model the world as having features that it in fact lacks. Our location gives rise to an internal model which involves location-dependent properties like 'left of' and 'right of', which do not appear in a location-invariant global description (one may also consider *de se* knowledge in this connection (Lewis, 1979a)). Perhaps more relevantly, our self-conception as agents seems to force on us a model of the world in which human beings are radically different in kind than the 'passive' matter that we may exercise our agency upon, despite the fact that, qua physical objects, we are in fact no different in kind from the rest of the population of our universe. It is quite clear in this case that our internal models of ourselves need not represent accurately any objective feature of our physical constitution. The models we choose to represent the way the world appears to creatures like us therefore may look very different to the models we choose to represent what the world is like from a global perspective.

There is no sense in which a perspective is not physical, since of course the facts about the perspective of some creatures may be studied naturalistically, by the sciences of psychology and biology. But they also provide explanatory resources that merely attending to the external world could not provide. For instance, an explanation of why some creature adopts a particular internal model need not appeal solely to the world that is being modelled, but may also appeal to the perspectival facts about the creatures who model the world in that way. It

is quite clear, I hope, that if the perspective of some creature is stable enough, it may well explain the persistence of some internal way of modelling reality without requiring that the model accurately represent the external world. Since the characterising perspective of an agent isn't a matter of choice or even conscious awareness for the most part, an appeal to that perspective may justify the internal models of that agent despite the fact that the they do not represent the external world. *A fortiori*, an appeal to perspective might justify the making of utterances on the basis of an internal model that do not represent the external world. And this, of course, is precisely what the minimalist wants.

Of course, there are non-minimalist (i.e. representationalist) ways to interpret this very same data. We may suggest that the linguistic practice correctly represents the internal model, not the external world. Or we could suggest that the practice aims to represent the external world, but fails. Both of these representationalist strategies fail to capture the distinctive role that the linguistic practice might play, because both fail to explain why the practice might persist as a successful strategy for these creatures to use in dealing with their external environment—the former cannot explain the external application, the second cannot explain the success. The non-representational minimalist interpretation, by contrast, explains both. A way of modelling reality that might be forced upon some creatures in virtue of the kind of creatures they are can be successful and persist if it is adaptive and conduces to survival. But not every adaptive trait must be explained as representing a response to some particular aspect of a creature's environment. So a practice that is explained by the nature of some creature that evolved under selective pressure can be justified without requiring that the practice is also explained by the fact that it accurately represents reality. This is particularly apparent in the case of practices that, if they represent at all, surely don't seem to represent anything that could have been exerting selective pressure. Numerical language and modal language are two prominent examples of manifestly successful practices that, if representational, represent apparently causally inert objects that could have exerted no selective pressure. As such, it can be no explanation of the success of such practices to appeal to representation. An alternative, non-representational account of the survival value of talking *as if* there were numbers or alternate possibilities is, while not without difficulties, at least some kind of

explanation (Rosen, 1990; Stalnaker, 1984b; Yablo, 2005).

On the minimalist picture, the justification for some linguistic practice, say the practice of describing ourselves as deliberative agents or the practice of number talk lies in the fact that those practices are conducive to success in creatures like us. Just how they do so is left open, though it is clear that the practice must somehow lead to predictive and explanatory strategies that facilitate our engagement with the external world. The fact that we typically engage in these practices as if they were fully representational is then explained by noting that it would undermine those strategies to explicitly regard them as false, involving us in some kind of pragmatic contradiction. However, in principle we could appeal to a naturalistic model of our behaviour, and recognise that our internal model may not be the best theory if our goal were only perfectly accurate representation with no other pragmatic considerations.

By endorsing the legitimacy of these perspective-dependent practices, we do not thereby automatically regard the objects of thought and reference in those practices as really existing in any sense. Insofar as these practices exist at all, they depend on the properties and relations of fundamental physics allowing for the existence of creatures whose constitution explains why they adopt the practices in question. It remains open to the minimalist to regard the explanatory strategies that a particular perspective-dependent practice makes available as on a par with the explanatory strategies of fundamental physics in some sense. It certainly seems to be true that these perspective-dependent practices have some degree of autonomy, since they may well persist whether or not we regard them as representing the reality they model (for example, even if there were a Platonic realm of numbers, it would play no explanatory role in the origin or structure of number discourse). We may thus get some limited devolution of ontological commitment from fundamental theories to less fundamental, perspective-dependent theories; yet in the end I doubt, and do not need to defend for my purposes, the idea that somehow this ontological commitment is to be taken completely seriously.

§4.2 THE CAUSAL PERSPECTIVE: COUNTERFACTUALS AND DELIBERATION

In §2, I identified a class of models that I take to characterise the core semantic aspects of causal language. Having just argued that we can sometimes give a minimalist justification for linguistic practices that need only appeal to facts about the users of those practices (§4.1), I now wish to argue that our causal language is itself a good candidate for a minimalist justification.

We begin with the obvious observation that we are *agents*. This involves having goals and projects, the capacities to accomplish some of these and the ability to deliberate about how to most effectively achieve them. It is clearly not an option for us to somehow abandon our agential status and refuse to deliberate, to act in some way as a mere object of physics. Thus our deliberative behaviour is practically inescapable. But here we may develop an observation of Ramsey's:

> ... from the situation when we are deliberating seems to arise the general difference of cause and effect. We are then engaged not on disinterested knowledge or classification... but on tracing the different consequences of our possible actions... (Ramsey, 1929: 158)

This suggests that the justification for our causal language might come from our deliberative practices, and not from any representation of objective reality. The basic feature of deliberation that is relevant here seems to be the importance of hypothetical reasoning: making a (counterfactual) supposition and tracing what might follow from it. In some sense, entertaining such counterfactuals is constitutive of rational deliberation. When considering what to do, we should consider the possible outcomes that might ensue given our act and then weight them by their how likely we regard them as being and by how much we desire that they come about. The theory of causation I've sketched has at its core this role in hypothetical reasoning (look, for instance, at how the conditional probability distribution induced by an intervention on a single variable corresponds to the likelihood of events conditional on actions); little surprise, then, that the practical necessity of deliberating leads on to the minimalist justification of causal talk, for at base they amount to the same thing.

What is important when deliberating is that we have a relatively robust and simple means for judging which events are dependent on one another, to facilitate

judgements about when an action we might undertake will be effective in bringing about our desired ends. We also wish to have a reliable and simple means of determining whether our activities will have further collateral consequences not themselves the intended goal of our activity. It should be clear that the account in §2 of causal networks of variables linked by counterfactual dependencies satisfy both of these desiderata. In virtue of these features, causal claims give us a quick and accurate way of accounting for why things happened the way they did: patterns of causal dependence summarise the relevant features of a situation that explain its overall character, as well as any event of particular interest.

Further support for the particular model of causation that I sketched above comes when we observe that our epistemic access to the external world is inherently limited to a local area. We are not able to perceive most of the possible variations in the antecedent physical states that might have some impact on the subsequent course of events, so it is important to rely on judgements of dependence that are relatively robust across a wide range of possible variations in background circumstances. Counterfactual dependence of some event on another holds only when the dependence is robustly invariant: otherwise it would not be the case that the consequent held in every situation where the antecedent held. If the events held to be counterfactually dependent on one another are relatively coarse-grained, then minor differences of detail will not be of great importance for figuring out what will come to pass.

It is crucial, then, that causal claims have some modal element, and cannot be reduced to merely actual physical connections. Robustness of a dependence relationship only makes sense if we can detect how that relation is stable under variations in the situation. This has further consequences: for example, when we make a request for an explanation, part of what we do is ask an implicitly general question: what was most responsible for the outcome, such that if we wished to reproduce it we would focus our efforts on that aspect? We rarely, if ever, request or give an explanation that accounts for an event in isolation and in such a way that the explanation doesn't generalise to similar cases. It turns out, therefore, that the giving of explanations depends in large part on the causal facts cited in an explanation being modally robust. Without robustness, the generalisability of causal explanations will fail. Without generalisable explanations, reasonable delibera-

tion seems impossible. The conclusion is obvious: deliberation is not optional for creatures like us, and the only way we could undertake it is if we model reality along the lines of the causal modelling framework of §2. Our causal talk, insofar as it is accurately modelled by that framework, is thus minimalistically justified: it is a framework that we must adopt given our nature, although that adoption does not commit us to there being any objective causal or counterfactual relations that our models represent.

As it stands the deliberative aspect of our agency seems like the most plausible feature of our perspective to ground the utility of causal models, and the above account of causation neatly latches onto this ground. Physical connection models of causal language (Dowe, 2000) do not have a close connection to deliberation, largely because it is implausible that modal claims may be determined by purely actual facts about causal influence. Given this, and Russellian worries about whether any candidate physical connection can play the causal role, the prospects for non-counterfactual theories of causation seem dim.

§4.3 THE INDISPENSABILITY OF CAUSATION: RESPONDING TO THE EXCLUSION ARGUMENT

At this point, the indispensability of causation starts to become apparent. Our deliberations require a modally robust relation between the action and the outcome. Causal relations fit this bill precisely. But the relation of global determination provided by fundamental physics is not robust, because in order to generalise from any one global state we need to abstract away some of the irrelevant details. But as we saw earlier, any abstracting away would bring about the failure of the determination relation, because precisely in the details abstracted away might lurk a potential event that would interfere with the situation supposedly determined by the parts of the state not abstracted away. The question is moot in any case, because without a pre-existing modal/causal concept fixing which parts of the state are relevant or irrelevant to the outcome in question, there is no sense to be made of abstracting away the unimportant details. Every detail is important; any abstraction trivialises the explanation. It is the very precision and detail of the explanation in terms of fundamental physics which means it is spectacularly poor at capturing the broad outlines of a situation. On the contrary, every detail matters

and none matter more than others, so any variation in even the most minor of respects leaves us unable to apply the lesson learned in the original situation (Eagle, 2004: 395).

This is precisely where causal explanations do well, and is precisely a reason to think that fundamental explanations do not trump causal explanations, contrary to Russell's assumption and the assumption of the causal exclusion argument. Of course, every physical event has some perfectly accurate explanation in terms of fundamental physics that describes with minute precision the actual circumstances of the event. But this emphasis on the details of the actual situation precludes the generalisations that every good explanation asks for: how can we apply the lesson of this situation elsewhere? In its emphasis on the actual, fundamental physics fails to give a reasonable answer to this question. The modal aspects of causation, particularly the counterfactuals that underlie true causal claims on my account, precisely answer this question concerning generalisations. Of course, even fundamental physics provides alternative models, and hence can ground some modal claims. But the only counterfactuals the most fundamental theories give us the resources to evaluate are those of the form 'If the global state were Γ, the subsequent global state after t elapsed would be Γ_t', and these are by no means the only nor even particularly important counterfactuals that we are concerned to evaluate when deliberating. We must recognise that, as we want robust dependence relations between salient local events, causal language does a much better job of describing and managing them than the highly extrinsic and gerrymandered relations that fundamental physics provides. It would be impossible to give up causal language without also abandoning our status as agents.

This emphasis on the incapacity of fundamental theory to deal with any but the most basic modal claims actually made no direct appeal to causation. Hence the objection raised by Russell would generalise to any alternate possibility not directly represented as a fundamental global feature of some other model of the fundamental theory. Any creature, therefore, that was concerned to represent or discuss alternate possibilities would be well served to avoid the conclusion of the causal exclusion argument—for soon on its heels would follow a 'chance exclusion problem' and a 'possibility exclusion problem'. If this were to happen, we would be in serious trouble. I've argued that our status as deliberating agents

depends on our being able to reason hypothetically and consider alternative situations. But there are arguments that I find extremely plausible that suggest that almost every feature of our thinking involves consideration of alternate possibilities. I am thinking here of Stalnaker's view that representation and inquiry are the fundamental practices in virtue of which we count as agents or thinkers at all (Stalnaker, 1984a). Representation requires consideration of alternative possibilities in a synchronic fashion, dividing the class of possibilities into those compatible with the current situation and those incompatible, and representing the actual world as a member of the first class. Inquiry requires consideration of alternative possibilities in a diachronic fashion, as incoming evidence and reasoning narrows the class of compatible possibilities over time. Both of these practices involve, and would be unrecognisable without, alternate possibilities. Insofar as modal notions are so central, we should all wish to adopt the minimalist defence of causal models against Russell's exclusion argument.

§5 Conclusion

The minimalist position with respect to causation is a third option between realism and eliminativism. Price (2004) characterises his minimalism as *republican*, by analogy with the political system, because it sees the source of justification for causal claims as being neither in the world, nor nowhere at all, but rather in ourselves and in the perspective-dependent practices we endorse. In some cases, inescapable features of ourselves, combined with practices it would be inconceivable to abandon, demand an inescapable commitment of some kind to practices that do not represent fundamental reality.

So it is with causation. Though Russell's arguments unfortunately preclude a realist reduction of causation to a deterministic fundamental physics (still less to an indeterministic physics), that does not mean that causation must be excluded from our conceptual economy. Once we recognise that we are limited agents, concerned with characterising effective and robust interventions on systems we care about, causal explanations and models have decided advantages over the explanations and models constructed using only the relations and entities explicitly found in fundamental physics. This is especially apparent once we realise that causation

stands or falls with other essential modal notions. No threat to the ontological pre-eminence of fundamental physics need follow from a commitment to the essential legitimacy of causation for agents like ourselves.

The minimalist picture fits naturally with a (hermeneutic) *fictionalist* interpretation of the discourse in question (Kalderon, 2005). This would involve regarding causal language as semantically continuous with the rest of language, but that the utterance of declarative sentences about causal relations does not have the force of an assertion that the semantic content of those sentences is literally true. (It may be that they are to be understood as assertions of something else, or not assertions at all.) The analogy holds however we understand these 'quasi-assertions': utterances within a given perspective, just as utterances in a fiction, are to be interpreted semantically as if the content of the fiction were true, but not as representing that the content obtains actually. When giving a naturalistic model of creatures as users of a particular theory, what we do is answer the question why these creatures should find it congenial or inescapable to adopt the fiction in question. I tried to answer that question for causation above (§4.2).

Though I myself prefer a fictionalist theory of causal discourse, nothing in what I have argued above relies on a fictionalist position. The narrow minimalist point required to avoid Russell's impossible conclusion is that causal explanations are practical necessities for agents like us, which provides a reason for keeping causation which is more than powerful enough to overmatch whatever reasons Russell adduces to get rid of causation. How we should then go on to understand the ontological status of causal relations—whether as having 'derivative reality' (Norton, 2003: §4), or being real but irreducible entities, or taking a fictionalist stance—is a different and much larger question that I cannot answer here.[23]

20 March 2007.

[23]Thanks to audiences at Princeton and at the *Causal Republicanism* conference, Centre for Time, University of Sydney. Particular thanks to Toby Handfield and Jeff Speaks; thanks also to Helen Beebee, Paul Benacerraf, Karen Bennett, John Burgess, Adam Elga, Mathias Frisch, Jason Grossman, Hans Halvorson, Gil Harman, Chris Hitchcock, Graham McDonald, Lizzie Maughan, Daniel Nolan, Huw Price, Gill Russell, Mark Schroeder, Brett Sherman, Charles Twardy, and an anonymous referee for Oxford University Press.

References

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

Bennett, K. (2003). 'Why the Exclusion Problem Seems Intractable, and How, Just Maybe, to Tract It'. *Noûs*, 37: 471–97.

Butterfield, J. (1992). 'Bell's Theorem: What it Takes'. *British Journal for the Philosophy of Science*, 43: 41–83.

Cartwright, N. (1979). 'Causal Laws and Effective Strategies'. *Noûs*, 13: 419–37.

Cartwright, N. (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*, 53: 411–53.

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

Eagle, A. (2004). 'Twenty-One Arguments Against Propensity Analyses of Probability'. *Erkenntnis*, 60: 371–416.

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

Field, H. (1980). *Science Without Numbers*. Princeton, NJ: Princeton University Press.

Field, H. (2003). 'Causation in a Physical World', in M. J. Loux & D. Zimmerman (Eds). *Oxford Handbook of Metaphysics*. Oxford: Oxford University Press, 435–60.

Hausman, D. M., & Woodward, J. (1999). 'Independence, Invariance and the Causal Markov Condition'. *British Journal for the Philosophy of Science*, 50: 521–83.

Hitchcock, C. (1995). 'Salmon on Explanatory Relevance'. *Philosophy of Science*, 62: 304–20.

Hitchcock, C. (1996). 'The Role of Contrast in Causal and Explanatory Claims'. *Synthese*, 107: 395–419.

Hitchcock, C. (2001). 'The Intransitivity of Causation Revealed in Equations and Graphs'. *Journal of Philosophy*, 98: 273–99.

Hitchcock, C. (2003). 'Of Humean Bondage'. *British Journal for the Philosophy of Science*, 54: 1–25.

Johnston, M. (1992). ‘Constitution is Not Identity’. *Mind*, 101: 89–105.

Kalderon, M. E. (Ed). (2005). *Fictionalism in Metaphysics*. Oxford: Oxford University Press.

Lange, M. (2002). *An Introduction to the Philosophy of Physics: Locality, Fields, Energy and Mass*. Oxford: Blackwell.

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

Lewis, D. (1973a). ‘Causation’, in D. Lewis (1986a). *Philosophical Papers*, vol. 2. Oxford: Oxford University Press, 159–213.

Lewis, D. (1973b). *Counterfactuals*. Oxford: Blackwell.

Lewis, D. (1979a). ‘Attitudes *De Dicto* and *De Se*’, in D. Lewis (1983). *Philosophical Papers*, vol. 1. Oxford: Oxford University Press, 133–59.

Lewis, D. (1979b). ‘Counterfactual Dependence and Time’s Arrow’, in D. Lewis (1986a). *Philosophical Papers*, vol. 2. Oxford: Oxford University Press, 32–66.

Lewis, D. (1979b). ‘Causal Explanation’, in D. Lewis (1986a). *Philosophical Papers*, vol. 2. Oxford: Oxford University Press, 214–40.

Lewis, D. (2000). ‘Causation as Influence’. *Journal of Philosophy*, 97: 182–97.

Menzies, P. (this volume). ‘Causation in Context’.

Norton, J. D. (2003). ‘Causation as Folk Science’. *Philosophers’ Imprint*, 3, URL www.philosophersimprint.org/003004/ [This volume]

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

Price, H. (1992). ‘Agency and Causal Asymmetry’. *Mind*, 101: 501–20.

Price, H. (2001). ‘Causation in the Special Sciences: the Case for Pragmatism’, in M. C. Galavotti, P. Suppes, & D. Costantini (Eds). *Stochastic Causality*. Stanford: CSLI Publications, 103–21.

Price, H. (2004). ‘Models and Modals’. In D. Gillies (Ed.). *Laws and Models in Science*. London: King’s College Publications, 49–69.

Ramsey, F. P. (1929). ‘General Propositions and Causality’, in F. P. Ramsey (1990). *Philosophical Papers*. Cambridge, UK: Cambridge University Press, 145–63.

Rosen, G. (1990). ‘Modal Fictionalism’. *Mind*, 99: 327–54.

Russell, B. (1913). ‘On the Notion of Cause’, in B. Russell (1963). *Mysticism and Logic*. London: George Allen and Unwin, 132–51.

Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction and Search*. Cambridge, MA: MIT Press, 2 ed.

Stalnaker, R. C. (1968). ‘A Theory of Conditionals’, in N. Rescher (Ed.). *Studies in Logical Theory*. Oxford: Blackwell, 98–112.

Stalnaker, R. C. (1984a). *Inquiry*. Cambridge, MA: MIT Press.

Stalnaker, R. C. (1984b). ‘Possible Worlds’ in Stalnaker (2003). *Ways a World Might Be*. Oxford: Oxford University Press, 25–39.

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

Woodward, J. (2001). ‘Probabilistic Causality, Direct Causes and Counterfactual Dependence’, in M. C. Galavotti, P. Suppes, & D. Costantini (Eds). *Stochastic Causality*. Stanford: CSLI Publications, 39–63.

Woodward, J. (2003). *Making Things Happen*. Oxford: Oxford University Press.

Yablo, S. (2005). ‘The Myth of the Seven’ in Kalderon (2005), 88–115.