ANALYSING CAUSATION
by
JENNIFER MARGARET MORGAN

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Department of Philosophy
School of Philosophy, Theology and Religion
College of Arts and Law
The University of Birmingham
July 2016

UNIVERSITY OF BIRMINGHAM

ii

This thesis will survey several prominent approaches to analysing causation, discuss their differences and similarities, and look at a number of problems which are common to all of them. I will be arguing for the following claims about how we should approach the process of analysing causation. Firstly, I will be arguing that a reductive analysis is desirable, since if we can reductively analyse causation in terms of something empirically accessible, we can explain how it is possible to know anything about causation. I will argue that to reductively analyse causation is to find out what kind of facts ground causal facts. Secondly, I will argue, following Hall and Strevens, that there are two kinds of causation, causal difference making and causal influence. This two-tiered approach explains the cases where we are tempted to ascribe conflicting characteristics to our concept of causation. Thirdly, I will argue that causal influence grounds causal difference making and that it does so necessarily. That the grounding relation holds necessarily is important for defending the two-tiered approach against the objection that it would yield a disjunctive account.

iv

Acknowledgements

First of all, I'd like to thank my supervisors, Nikk Effingham and Al Wilson. Your advice, suggestions, patience and persistence in talking me out of quitting, repeatedly, have been just amazing.

I am grateful to my family, in particular my parents, David and Hannah Stone, and to the memory of my grandparents Brian and Margaret Williamson, without whose generous support I could never have undertaken this degree. And thank you, Dad, for your technical help with the formatting of citations; you have now helped me with my homework for the last time, ready for my coming of age at 33.

My co-workers in my day job have been very kind, in some cases covering and giving me time off to work on this thesis. I am grateful for their thoughtfulness and forbearance, whilst I tried to labour in my day, vocation and degree all at the same time.

My friends have been wonderful. In particular I am grateful to Jake Wood for saying that he was looking forward to reading this thesis, although not half as grateful as I am to him for being someone who wouldn't have said it if he didn't mean it and whom I could therefore, to my surprise, believe. And I am endlessly grateful to my best of all possible friends, Nicky McConkey Wood, for her support, pep talks, thoughtful questions, kind suggestions and persistent belief that I was going to finish this one day.

Finally and most importantly, I am grateful to my husband, Bob Morgan, for his love, material and emotional support, kindness and belief in me.

Contents

IV   Two kinds of causation   215

Introduction

This thesis will argue for the following claims:

1. To analyse causation is to find out what kind of facts ground causal facts. I will argue that a reductive analysis is desirable, since if we can reductively analyse causation in terms of something empirically accessible, we can explain how it is possible to know anything about causation.

2. There are two kinds of causation, causal difference making and causal influence. This two-tiered approach explains the cases where we are tempted to ascribe conflicting characteristics to our concept of causation.

3. Causal influence necessarily grounds causal difference making. That the grounding relation holds necessarily is important for defending the two-tiered approach against the objection that it would yield a disjunctive account.

The subject matter of this thesis is the question of how and why we should analyse causation. Causation is a ubiquitous concept, in philosophy, science and everyday life. The analysis of causation, like that of mind, knowledge, language and other philosophically interesting concepts, has received a lot of philosophical attention. In this introduction I will briefly discuss a few features that our concept of causation is usually taken to have. I will then give a sketch of an explanation of why analysing causation is not a straightforward task, and why comparing different analyses of causation can be like comparing apples and oranges. I will explain how these accounts differ not only

in the analysis of causation which they offer, but in their aims, background assumptions and methodology. I will then give an outline of my thesis and the claims I will be defending in it.

0.1   Characterising causation

The following are commonly taken to be features of causation. Many of them are disputed, but together they do give us a sketch of the target of our analysis. In the table below, I have grouped views into the received and alternative views. Five of these characteristics, Transitivity, Locality, Intrinsicness, Dependence and Absences, have been suggested by Hall (2004, p. 225) , while others have been mentioned elsewhere in the literature. This table provides a rough idea of what we mean by causation, and what aspects of it are contentious.

Characteristic	Received view	Alternative view(s)
A relation	Causation is a two place relation, just possibly a four place relation.	Causation is not a relation. It is an operator. (Mellor 1995)
Relata distinctness	Causation takes place between distinct relata (Fine 2012)	Inertia is a form of causation. (Dowe 2000, p. 52).

Transitivity	Causation is a transitive relation. See both early Lewis (1973a) and late Lewis (2004a)	Causation may not always be transitive. (McDermott 1995) (Kvart 1991) (Hitchcock 2001). See Hall (2004, p. 183) for discussion.
Intrinsicness	Causation occurs between the causal relata in virtue of their intrinsic properties. Changes in something extrinsic to the situation will not change whether or not causation happens. (Hall 2004), (Menzies 1996)	Causation can be extrinsic (regularity theory) (Beebee 2006), (Hempel 1965), (Baumgartner 2008)
Locality	There is no action at a distance, without something to mediate. (Hall 2004)	Some understandings of gravity and quantum entanglement imply action at a distance.

Dependence	Causation is intimately related to counterfactual dependence. (Hume 2000 [1739]), (Lewis 1973a), (Lewis 2004a), (Hall 2004)	Counterfactual analyses of causation are a mistake, or at best only characterise our concept of causation rather than what causation is in reality. (Dowe 2000), (Fair 1979), (Salmon 1984).
Absences	Absences can be causes and effects.	Absences cannot be causes and effects; if we rephrase causal claims about absences we will see this (Dowe 2000, p. 125).
Decision making	Knowledge of what causes what should play a key part in our decision making. (Causal decision theory.)	Knowledge of causes should not play a part in our decision making. (Evidential decision theory.)

Determinism	Causation is deterministic. One thing necessitates another.	Causation can be indeterministic. Certainly at the quantum level, causation occurs where the cause does not necessitate the effect. (Dowe 2000), (Paul and Hall 2013, p. 63) [1]
Manipulation	Knowledge of causes allows us to manipulate situations, but cause cannot be defined in terms of manipulations.	Causation can be defined in terms of what can be used to manipulate what. (Collingwood 1998), (Von Wright 1971), (Gasking 1955), (Woodward 2003) (Menzies and Price 1993).

---

[1]Note that most philosophers agree with this, but I suspect most of the folk regard causation as deterministic, hence my categorising determinism as the received view and indeterminism as the alternative view.

Biff	There is a relation called biff which is present when there are physical pushes and pulls, and exchanges of energy and momentum. This is closely linked to causation. (Lewis 2004b, p. 283)	This biff relation *is* causation. (Dowe 2000)
Between particulars	Causation is fundamentally a relation between particulars. General claims about causation depend for their truth on claims about singular causation. (Mellor 1995), (Ducasse 1969), (Dowe 2000)	General causal claims are more fundamental and the truth of singular causal claims depends on them. This is characteristic of regularity theories (Hume 2000 [1739]), (Psillos 2002). Alternatively, Sober (1984) argues singular (token) and general (type) causation are two separate concepts, but does not take a stance on which is more fundamental.

Asymmetry	Causation is asymmetric, and not just temporally. Causation runs from cause to effect and this isn't just because causes come first. Backwards in time causation is possible in principle. (Lewis 1979).	The asymmetry of causation depends on the asymmetry of time (ie causes are different from effects because causes come first). See Dowe (2000, p. 50) for discussion.
Fundamentality	Causation is not fundamental. Russell (1912) famously argued this, and went on to say that for this reason we should drop the concept of causation from all serious scientific and philosophical conversations. Most philosophers would now take the view that although causation is derivative, it is still important in both science and philosophy.	Causation is fundamental. Cartwright (2004) has a view of causation as many distinct and irreducible causal relations which implies that causation, or rather the many varieties of it, are fundamental. Frisch (2005) argues that causal notions have a place in fundamental physics.

Mysterious	Causation is basically mysterious. It is hard to understand what it is, and hard to come by reliable causal knowledge. (Hume 2000 [1739]).	Causation is something we encounter everyday. When we say we don't see causal relations, we are just setting the bar for 'seeing' something ridiculously high. (Anscombe 1971) Analysing causation is relatively straightforward (Ducasse 1969).

0.2 Why is analysis a complicated task?

Let's take one example approach to analysing causation and show how philosophers have disputed various aspects of it. One approach to analysing causation is to ask how to complete the sentence 'c causes e if and only if...'. We might take some way of completing that sentence (for example 'if c hadn't happened then e would not have happened either') which doesn't itself involve the word 'cause' or other causal language. We would then search for counterexamples to it. That is, we would look for examples of scenarios where we want to say that one thing caused another, but according to our analysis, it didn't. Or, cases where we would naturally say that c did not cause e, but according to our analysis, it did. If we find a counterexample,

we will need to adapt our analysis so that this counterexample is no longer a problem: so that all the 'right' cases are counted as causation. This is known as the method of cases, or sometimes as conceptual analysis.

This sounds like a pretty straightforward approach. If all attempts to analyse causation went along these lines we could simply examine each one in turn and see which had the fewest counterexamples. We could then assess which of the analyses of causation was the winner by these standards.

However, the situation is more complicated. To begin with, there may be some apparent counterexamples, which we initially see as making trouble for our account of causation, that we refuse to accept as counterexamples at all. Instead of giving up or modifying our analysis of causation, we might claim that the example is of something which seems like causation, but according to our analysis, really isn't a case of causation. Our everyday intuitions are wrong about that case. This brings out a question about the analysis of causation (and analysis in general): to what extent should our analysis be revisionary? Are we looking for a story which perfectly reflects our everyday intuitions? Or are we looking for something which we could use to correct our judgements and set us straight in problem cases?

This is not the only contentious issue. Almost every feature of the approach to analysis which I mention above has been challenged by someone. It is based on our intuitions about cases and thus about the concept of causation. To some philosophers such as Dowe (2000, p. 2) this isn't what we want from an analysis of causation; they want to get an analysis of what causation is in reality, in the world, not what our concept of causation is. To others, we need a two stage process to capture something which is both what

we mean by causation, but causation as it takes place in reality and not our concept of it; I will discuss this in my chapter on the Canberra Plan.

Other philosophers such as Woodward (2003) do not agree that we need to analyse causation in a non-circular manner. That is, we are not trying to define cause entirely in non-causal terms. It is acceptable to produce a circular analysis as long as the analysis is informatively circular. They argue that the important project is not to reduce causation to something else but to clarify what the different causal concepts are, the links between them and the links between our understanding of causation and the methods we use to find out whether causation is going on in the world. We might think that the concept of causation is fundamental – that it cannot be defined in terms of anything else – or that reducing it to something more fundamental will not give us the kind of insight into it that we want.

It is not uncommon for philosophers to suggest that there is more than one causal concept. Sometimes they will argue that another philosopher's analysis of causation actually captures some related concept, rather than causation itself; Lewis suggests that the analysis of causation due to Fair (1979) is really an analysis of 'biff' (Lewis 2004b, p. 283). Others argue that we need more than one analysis of causation: we have two concepts of causation which we frequently confuse cite.

Every analysis of causation has some counterexamples which make trouble for it; no analysis is widely accepted. What counts as a successful analysis will also depend on what that analysis is for, so as well as the question of how we analyse causation, we need to look at why.

0.3 What in particular is challenging about analysing causation?

So far the problems we have listed are part and parcel of analysing any philosophically interesting notion. Each concept that's of interest to philosophers – knowledge, truth, the good, personhood and so on – can be expected to come with it's own set of *unique* problems as well. Causation has its share of these.

One issue is that causation, unlike many other philosophically interesting concepts, is a relation. (At least it is usually taken to be so.) It relates two other things. So philosophers of causation have an additional question to answer: what are the relata? No analysis of causation will be complete without an answer to this. I will discuss this issue in chapter two below.

This comes with an additional problem: the worry about missing relata (Lewis 2004b). There are some cases where it seems that one or more of the things related by causation is an absence. This presents all kinds of problems for theories of causation, which I will discuss in more depth in chapter three. While some of these problems may be endemic when it comes to relations (for example, falling into a hole[2].), they are particularly acute in the case of causation because it is so central to science, philosophy and every aspect of the day to day world. It is difficult to deny or explain away cases of absences causation which occur every day.

The only case which seems at all similar is that of grounding. We might think that the absence of a hippo in the lake on the University of Nottingham

---

[2]This example is due to Liebesman (2011)

campus grounds the truth of the claim 'there's no hippo in the lake on campus' (this example is due to Jago et al. (2012)). The relation of grounding in that case would then have a relatum which was an absence. So causation might not be entirely alone in having this problem. Nevertheless, it is not a widely shared issue.

Another unique feature of causation is the role it plays in prediction and decision making. Certainly knowledge, goodness, truth and so on can be important concepts to us when we are making decisions. But they are not important in quite the same way. Causation affects, at least according to some, the very way in which we partition the decision space. Other concepts merely relate to what value we might put on each possible option.

For example, take the case known as Newcomb's Problem (Nozick 1969). In this case, we have two boxes: the first is opaque and contains either £10k or nothing. The other is transparent, and contains £10. The choice is whether to take one box, or two boxes. The catch is that an infallible predictor knows which box you are going to take, and has put £10k in the opaque box only if you are going to take just the opaque box. If you are going to take both boxes, then the predictor will leave the opaque box empty. There are two camps on how to respond to this puzzle: one-boxers and two-boxers. For the one-boxers, it makes sense to choose to take one box because that means the infallible predictor will have put £10k in that box, whereas if you took both you would only end up with £10 as the infallible predictor would have put nothing in the opaque box. However, two-boxers would take both boxes because, after all, the infallible predictor has already decided: he has either put £10k in that box or nothing. So, we should take both boxes:

even if there is nothing in the opaque box we will at least get £10. In any case, two-boxers argue, we cannot make there be £10k in the opaque box by taking only that box: it has already been decided.

In this case we can see how causal considerations play a profound role in decision making. For two-boxers, causal considerations are important in making the decision; for one-boxers they are not. For two-boxers, the infallible predictor has already made his decision about what to put in the opaque box. Nothing can be done to change that now. We cannot cause something to have happened differently in the past. This causal consideration is what motivates the two-boxers decision. Whereas the one-boxer knows that if they are going to take just the opaque box, the infallible predictor will have put £10k in it. The fact that what they do now cannot cause the infallible predictor's choice does not enter into it. The point is here that we not only have a disagreement about what the best option is, but about what the available options really are.

This raises the stakes for analysing causation. On the one hand, it seems like we want to say causation is mind-independent, but on the other it is very deeply embedded in human practices of decision making, prediction and explanation.

Another unique issue with causation is its relation to counterfactuals. Although strictly non-counterfactual accounts have been offered, there is still something intuitively right about idea that one event causes another only when, if the first event hadn't happened, the second would not have happened either. Counterfactuals are notoriously metaphysically troublesome, since they are claims which are supposed to be objective, and yet are not about

the actual world. So coming up with a metaphysically respectable account of causation is hindered by its link to counterfactuals.

0.4   Thesis outline

I will begin with some issues about how we understanding causation are partly separate from what analysis of causation we choose to adopt. First, I will discuss what analysis is. I will argue that when we are analysing causation we are claiming that a grounding relation exists between causal facts and facts involving the analysans.

Next, I will discuss the issue of the causal relata. Causation is commonly taken to be a relation. If that is so, an important question to ask about it is what it relates: events, facts, objects, aspects, or something else? There is also the issue of causation by and of absences, which deserves to be tackled separately from any chapter on any particular analysis of causation, since it is a problem for all of them. These chapters will introduce issues which we can keep in mind throughout the thesis.

Next, I will look at three prominent approaches to analysing causation, all of which are reductive analyses. In other ways these approaches are very different however: Lewis's account is counterfactual, Dowe's is empirical, and the Canberra Plan approach aims to find something in reality which satisfies our folk concept of causation.

In part three, I will move on to discussing non-reductive approaches to analysing causation. A good example of this approach is Woodward's; I will look at his account in detail. We will look at what these non-reductive ap-

proaches can offer us, and why some philosophers find them more promising than reductive accounts. I will go on to consider two reasons why we might still seek a reductive account, despite the non-reductive accounts available: ideological parsimony and empiricism.

In part four, I will discuss causal pluralism: the possibility that there is more than one concept of causation and that we really need more than one analysis of causation. I agree with Hall's diagnosis as to why we struggle to analyse causation. Hall (2004) argues that there are several theses which we want to think are all true of causation, but they cannot all be true at once; this explains our problem with analysing causation. I will suggest though, following Strevens (2013), that facts about one kind of causation reduce to facts about other. One of our two kinds of causation is fundamental, and the other is grounded in it.

I will be arguing for the following claims. Firstly, a reductive account of causation is desirable. This is because a reductive account could allow us to reduce causation to something empirically accessible. That is, causation, as it stands, is not something we can observe. An analysis of causation could show how causation is grounded in something in reality which we can observe. This could, in turn, explain how it is possible that we can know anything about what causes what.

Secondly, we will need more than one account of causation. One account will be a process account, which gives us an idea of what causation is entirely in terms of conservation and exchange of conserved quantities; this is Dowe's analysis of causation. This account of causation, causal influence, provides us with a link between something we have empirical access to and

causation. It also explains why we regard causation as intrinsic, transitive and local. However, this account of causation fails to account adequately for causation by and of absences, nor for our intuition that there is a link between counterfactual dependence and causation. So a second account of causation is required, based on counterfactual dependence: causal difference making. I will argue that facts about causal influence ground facts about causal difference making.

Part I

The Causal Relation

Chapter 1

The analysis relation

1.1 Introduction

In this section I shall argue that analysis is the discovery of a grounding relation. That is if we analyse term a in terms of term b we are saying that facts about a are grounded in facts about b.

First, I will argue that the analysis relation is not one of identity. Identity is undirected, whereas the analysis relation has a direction; it is asymmetric between analysans and analysandum. I will then argue that the grounding relation is more suitable than identity here. I will also discuss what the target of our analysis of causation should be: that is, whether it should be the word / concept of causation or what causation is in reality. I will touch on some issues regarding our reasons for analysing which will be discussed in more depth in chapter 9.

1.2   What is analysis?

In the literature on analysing causation, the term analysis is used broadly. It encompasses any attempt to say what causation is, or to say what a cause is, or to characterise what it is for something, like a process, to be causal. This includes conceptual analysis, the Canberra Plan, empirical analysis and pragmatic analysis. The analysandum is not necessarily the one definitive causal relation, but may be a variety of causal concepts.

Some analyses of causation are not reductive; we will move on to looking at these in the next part of my thesis. Others are not supposed to be necessarily true. Dowe describes his approach as contingent analysis: his analysis tells us what causation is in this world, but in other possible worlds it could be something else. He does not think the contingency of his account stops it from being an analysis.

Some words such as 'accounts' and 'approach' are suspiciously similar to 'analysis', so a little disambiguation here may be useful. I will use the word 'account' interchangeably with 'analysis'. 'Approach' is a broader term covering the general type of analysis and method for analysing causation, as well as the analysis itself.

1.3   Is analysis identity?

Despite this diversity in what kind of approach can count as an analysis, there are still some positive comments we can make about what it is to analyse something. To begin with, if we say we can analyse causation in

terms of certain conditions, then some sort of analysis relation exists between causation and those conditions. So we can then ask questions about the nature of this analysis relation. In particular, is it the same as identity? This seems like an obvious possibility, since we are likely to set out our analysis of causation by saying "to cause something is to ...". Both Chalmers (1996) and Kim (2008) talk as though reduction is, or requires identity, even though this might mean setting reduction to one side in favour of the more useful concept of reductive explanation.

However, there are reasons to think that analysis is not the same as identity. Identity is undirected. If a is identical with b, there is no real sense in which one of those can be prior to and the foundation for the other one. Identity is reflexive: everything is identical with itself. Analysis, on the other hand, is not. We cannot analyse a concept using itself, perhaps because providing an analysis is supposed to explain and inform, and repetition does not achieve that. We do, however, want the relation of reductive analysis to be transitive: if I analyse a in terms of b and b in terms of c then a is ultimately reduced to c. For the analysis relation to be both irreflexive and transitive, we need it to be asymmetric.

This asymmetry fits well with the idea that analysis is a discovery. Certainly we can have a genuine discovery that two things are identical with each other ("Hesperus is Phosphorus"[1]) but in that case we have discovered just as much about each of the things which turned out to be identical. At least, discovering the identity directly tells us exactly as much about each of the things identified, leaving aside any inferences we may be able to make.

---

[1] To use a classic example due to Kripke (1980).

In the case of an analysis, we are supposed to have discovered more about the analysandum than about the terms into which we analyse it.

What else could analysis be apart from identity? We can look for clues in how we phrase analysis claims. We often write something like "*c* causes *e* if and only if *a*". This suggests that the relation between the analysis and analysandum is strict implication; if *a* holds (whatever those conditions are) then that implies that causation holds too. However, we would have to say that this is mutual strict implication; if causation holds, then that implies that analysis conditions *a* also hold. So this does not give us the directedness we want: the sense that in doing analysis we discover more about the analysandum than about the analysis conditions. Also, we can easily have cases of mutual strict implication which are not analysis: for example if $x + 2 = 4$ then $x + 3 = 5$; each of these claims entails the truth of the other, but neither is an analysis of the other. So analysis is not mutual strict implication: that is both too broad and too narrow.

In the next section I will go on to suggest that reductive analysis is the discovery of a grounding relation.

1.4   Grounding

So, analysis is very like identity. However, it is asymmetric in how it informs. That is, the analysis tells us more about the analysandum than the analysandum tells us about the analysis. It looks as though analysis, as a relation between analysandum and analysis terms, resembles grounding. Grounding is a metaphysical relation between objects or other phenomena, which re-

lates to explanation. While one form of explanation is causal, another form is grounding explanation: a glass is fragile in virtue of the arrangement of molecules which makes it up (Rosen 2010, p. 110). Fine (2012) writes that grounding is both 'explanatory' and 'determinative' and thus fills an important role in metaphysics.

Understanding analysis as the discovery of a grounding relation would provide us with a metaphysical understanding of it, just as if we had said that analysis is identity. It would be clear that we are talking about analysis of causation in the world, not the word 'causation'; I will discuss this in more detail later in this chapter. Grounding would have the additional advantages that it includes the directional element of analysis, as well as the explanatory element.

Fine (2012) argues that the notion of grounding in important in realist metaphysics. Realist or critical metaphysics is, according to Fine, the kind of metaphysics which is concerned with saying what is real, as opposed to naive or pre-critical metaphysics which is concerned with the natures of things. Grounding offers us a way of say how some things which clearly exist in some sense can be non-fundamental. Supervenience will not do the job, because it is not directed as grounding is. To say that A supervenes on B is to say that there is no A difference without B difference. But this works both ways. For example, that the mental supervenes on the physical is not really an answer to questions in the philosophy of mind. The notion of grounding is directed and so is more useful for expressing positions on this kind of question.

Fine (2012) suggests that grounding is a way of understanding the common idea of reduction or analysis. Rosen (2010, p. 122) also says something

like this. To Rosen, reduction or an analysis is a relation between propositions. 'Real definition' is between items (ie not the words which refer to them). Grounding, to Rosen, is similar to 'real definition' but unlike 'real definition' it is irreflexive, and between facts.

With this in mind, I suggest that when we say we have reduced, or reductively analysed, one concept in terms of another, we are claiming that facts about the analysans ground facts about the analysandum.

1.5   What is the target of an analysis of causation?

The question here is whether we are analysing the word 'causation', the concept *causation* or the mind-independent relation, causation.

One way of thinking of it is that we want to analyse the word 'causation'. The idea is that an analysis is supposed to give us a substitute for the word 'causation' or 'cause'; we can take those words out and swap our analysis in. This is what Ducasse (1969), for example, took himself to be doing.

If this sounds too much like dictionary writing, there is also the view that what we are analysing is the concept *causation*. This might be what is often meant by conceptual analysis. Dowe (2000, p. 2) compares conceptual analysis to empirical analysis, arguing that they can exist side by side. An empirical analysis is supposed to tell us what causation is in reality, while a conceptual analysis is supposed to tell us what our concept of causation is. There is a difference between a concept and a word. For example, the

word 'plus' is different from the concept of addition. Different languages will usually have different words, but often have the same concepts. In the case of causation, the causal concept is in play when we look at neuron diagrams for example, or structural equations, even though the word cause is nowhere in sight.

So there are three possibilities for what the target of our analysis might be: the word 'causation', the concept *causation* (conceptual analysis) and causation in reality (empirical analysis).

One way to decide what our target should be might be to look at our reasons for analysis.

1.6 Reasons for analysing

One important reason for analysing causation is the hope that it will shed some light on the epistemic and psychological side of things. There are two ways that an analysis of causation could help us with epistemic questions, such as when we are trying to discover if causation is taking place in a particular instance. The first and most obvious is that an analysis could function as a kind of checklist, to be used when we are trying to identify cases of causation. We might have a case where some event followed another and want to know if the first caused the second; we could work our way through our analysis of causation, to see whether all the conditions applied in this case. This would give us an answer about whether our case was causal or not.

This is not the main kind of epistemological help we are looking for how-

ever. We are able to identify cases of causation all the time, quite easily. If we are using a checklist which corresponds to an analysis of causation, we are not aware of it. Equally, there are cases of causation which are not at all easy to identify, but when scientists do identify them, they are fairly clearly not running down a philosophical checklist. The methods are more complicated, and contain a larger mathematical component. If there is an analysis-of-causation checklist, again it is implicit.

An analysis will not give us much epistemological help here, because it seems like we don't need it. Either we are getting along perfectly well without an analysis, or we already have one, and just have not made it explicit. If we do have an implicit analysis, then we might want to spell it out explicitly. This might well satisfy our curiosity about how our minds work, but it doesn't seem like it would be any epistemological help, since we are managing quite well now with an implicit analysis. One further possibility is that we might have an implicit but partial or partially correct analysis of causation, which we use when we make causal judgements. In that case, we might want to make explicit and then correct our analysis.

This seems the most plausible way in which an analysis could help us identify cases of causation. It is only really likely to be helpful if we suspect that we are getting causal judgements wrong, and that the reason why we are getting them wrong is that we are using the wrong implicit checklist. In that case, an analysis of how we currently use the word or the concept of causation, rather than causation itself, will not be helpful. If a definition accurately describes how a word is used, then that definition is correct. There is no scope for saying that we are using the word wrong and need a different

checklist. If our philosophical process of analysis is supposed to help us by correcting our checklist, then our checklist must point to something non-language dependent, which we have the potential to be wrong about.

The fact remains, however, that in really tricky cases where we do not know what causes something, we do not turn to a philosophical analysis to resolve the problem. I think the kind of epistemological help we can expect from an analysis of causation is mainly of a different kind. Apart from the questions about particular instances – 'how do we know causation is happening in this case?' – we also have the general question – 'how do we know anything about causation?' The question here is how we can know causal truths, since they seem different from facts which we can get from empirical evidence, and from a priori mathematical truths. As well as knowledge, we can also ask how we come to have beliefs about causation at all. This is the question Hume asks: where does our idea of causation come from?

Here's one way an analysis could be expected to help us with this question. If we assume that there are certain kinds of beliefs where it is unproblematic to know where they come from, but our beliefs about causation are somehow mysterious in origin, then an analysis of causation in terms of something unmysterious may explain how we can have beliefs about causation, depending on what the analysis relation is.

Here's another way of putting the same suggestion:

1. There is no mystery about how we acquire beliefs about some things. Call these things transparent things.

2. But how we acquire beliefs about causation is mysterious. So we can

describe causation as non-transparent.

3. We can analyse causation in terms of transparent things.

4. The analysis relation is such that if we analyse something non-transparent in terms of something transparent, the non-transparent thing becomes transparent to us.

The thought is this: with some things, we are perfectly clear about how I acquire my beliefs about them. For example, it seems comparatively obvious how I come by beliefs about objects I directly perceive. (I do not mean to suggest that perception is problem-free, but I want to distinguish the particular problems we have with some things we do not directly perceive.) It also seems relatively clear how we acquire beliefs about a priori facts about logic and mathematics. But causation seems to be problematic in a way that these are not.

Most writers, and famously Hume, claim that we do not see causation. We see the events that are the cause and effect, but we do not see that they are the cause and effect; we do not see the relation between them. Some writers such as Anscombe (1993) and Cartwright (2004) will insist that we do see causation, just as we see ordinary items and events.

How is it that we understand what causation is, and will point out certain events to be causes and effects? How do we use empirical evidence to justify these identifications of causes and effects? Obviously a detailed account of how we do this would need to involve a scientific methodology; the philosophical question is rather, how is it possible at all?

We can go through the different possible targets of an analysis of causation, and look at how much help each kind of analysis could be with these epistemological questions. Firstly, let's look at analysing the words 'causation' and 'cause'. Having a definition of 'cause' might help us see where the causes are in a basic sense. A definition will tell us what we mean by a word, so we know what to apply it to. For example, if we define 'marsupial' as 'a mammal whose young are born incompletely developed and who are nurtured in their mother's pouch', we know to pick out the marsupials as marsupials, but that birds are something else.

It may be helpful to have a definitional account of causation, since moving the goalposts is a poor, but common, strategy in an argument. But to know whether we are moving the goalposts, we need to know where they should have been in the first place. In other words, we want to know what causation is, so that we can know what it isn't. We would like to know in what cases we can legitimately say: 'what you're talking about is not causation'. There are some contentious aspects of the definition of cause which are likely to come up when we are making judgments about specific cases of possible causation, or more pressingly, specific cases of responsibility. One example would transitivity: if *a* causes *b* and *b* causes *c*, does that mean *a* causes *c*? Another contentious area might be indeterministic causation: what should we say about cases where the cause does not guarantee the effect, but we nevertheless want to call what is going on 'causation'? Providing an analysis has the potential to clear up these contentious areas, or at least make it clear what other commitments we may be signing up to when taking up a position on them.

Giving an analysis of the word 'causation' does not really help us with our epistemological questions about causation, however. In the case of 'marsupial', there is nothing mysterious about how we recognise what mammals have a pouch. Our definition of causation will be telling us how we use the word 'causation'. In order for it to illuminate how we come to recognise causes, it must define 'causation' in terms of something that we can recognise, and where we find nothing epistemologically mysterious about how we recognise it. Then we come to a dilemma. Either the definition of our word 'cause' is in terms of something unmysterious, in which case the question is why we have a word for those things: why it is important to us to pick them out. The analysis tells us that we should pick out certain events as causes, and we know how we recognise those events. But we still do not know why we think of those events as causal, beyond 'the language works like that; those events all get the causal label'. The other alternative is that the definition of 'cause' is in terms of something such that we still don't know how we know about it. In that case, the definition is of no help with our epistemological questions. The cases where we are trying to analyse the concept of causation if effectively the same. The concept is different from the word: after all, there will be different words for causation in different languages, and we may use the concept of cause without using the word, for example, it might be represented mathematically. Nevertheless, for our current purposes we can treat these two as the same.

Secondly, take the case where our analysis of causation is meant to be an analysis of causation in the world, not of a word or concept. In this case, analysing causation is supposed to be a discovery about the world. Saying

that this analysis would tell us "what causation is" suggests that the analysis relation is the same as identity, which seems premature and implausible. However, the phrase "what causation is" does capture the thought that the point is to find out something about causation itself, not our word or concept. It seems that a reductive account of this type might help us with our epistemological questions. If we find that causation can be analysed in terms of something which is epistemologically transparent to us, then we have explained how we can acquire and justify beliefs about causes. This does not suffer from the same problem as with the definition approach to analysis. We do not then need to ask why we use the word 'causation' for whatever it is we have analysed causation in terms of. The analysis of causation should explain what it is that everything causal has in common, beyond that we use the same word for it. When we come up with our analysis of causation in the world, we will have made a genuine discovery about the world and not about language. I will discuss the need for a reductive analysis in more detail in chapters 7 and 8 below.

Another reason to want an analysis of causation is to find out if it fits in with our other theoretical commitments. This follows on from the previous point: if we discover that the best analysis of causation is one according to which causation is mind or language dependent then this will not sit well with a view of science according to which it discovers truths about the mind-independent world.

1.7   What kind of analysis of causation should we look for?

In this section, I will provide a rough taxonomy of different kinds of analysis of causation.

An analysis of causation may be reductive or non-reductive. Reduction and analysis are sometimes used synonymously, which can make the idea of a non-reductive analysis confusing. However, a bit of disambiguation will clear this up. We can take analysis as a relation between two things, which is directed, explanatory in one direction, and something like identity (more on this later). We can analyse a in terms of b and then b in terms of c. We can talk of 'the analysis' in the sense of one of these relations, or use it to describe the whole chain of relations. In a reductive analysis, the chain of analysis relations stops somewhere, with something which does not itself need analysing. The reasons why we stop there might vary: it might be because we've achieved some kind of parsimony, or we might stop with something more empirically accessible. But at some point we reach a stage where no further analysis is needed.

In the case of a non-reductive analysis, we may analyse a in terms of b, in terms of c and c in terms of a. The circle may be more or less tight; there may be more of fewer steps to this analysis. This is, to all intents and purposes, a circular definition. However, some writers such as Woodward (2003) have defended non-reductive analysis and argued that it can be informative.

Non-reductive analysis can be understood as the same as reductive analysis, but whilst remaining neutral on the direction of reduction. I will discuss

this in more detail in part three. Another, more orthodox way of understanding non-reductive analysis is that it has a direction, but that the chain of analysis goes around in a circle. For example A is defined in terms of B, B is defined in terms of C whilst C is defined in terms of A. The value of such an analysis is normally taken to depend on how tight that circle is, and how helpful it is in illuminating the relationship between the concepts involved.

1.7.1 Pluralism and contingency

Our analysis could be univocal or pluralistic. In other words, we might have just one concept of causation, or more than one. We could say that c causes e iff some non-disjunctive set of conditions apply; this would be a univocal analysis. On the other hand, the conditions could be disjunctive: c causes e iff some set of conditions hold or some other set of conditions hold. There could be more than two disjuncts; in principle there is no limit. However, any disjunctive analysis of causation will have to come with some convincing story about why we call both of these disjuncts causation. If the analysis is disjunctive, why do we use the same word? Assuming it isn't a simple case of homonymy, such as mean (average) and mean (unkind), there must be some connection between the two concepts of causation. But this then raises another question: why isn't this connection something we can express in the form of a univocal analysis? Any pluralistic analysis is going to be pulled both ways: on the one hand, it has to include enough of a difference between the two concepts to justifify giving a pluralistic analysis rather than a univocal one, but at the same time, must explain what both of the concepts

have in common which makes them deserve to be called causation. I discuss pluralist analyses in more detail in part four.

The analysis relation is often understood as a necessary relation, so that whatever the true analysis aof causation is, it is necessarily true. However, in the literature, the word analysis is used to include contingent analyses as well, such as Dowe's empirical analysis (Dowe 2000). So whether an analysis should be necessary or contingent is up for grabs. A contingent analysis will have questions to answer, however, rather like a pluralistic analysis. A contingent analysis of causation may only hold true in our actual world. In the actual world, c cause e iff some set of conditions, while in another possible world c causes e iff some other set of conditions. So then we need to say why this other phenomenon, in the other possible world, should be called causation too. Why not say that our analysis of causation for our world was necessarily true after all, and whatever it is that goes on in the other world is similar, but not causation? The situation parallels the questions about pluralistic analyses.

The idea behind the Canberra Plan methodology (discussed in chapter 6) is that there are two parts to our analysis. First of all there is the conceptual part in which we work out the role that we are talking about when we use the concept in question. Then once we have established that, we look for what fills that role in reality; this is the empirical part of the analysis. This approach could serve as an answer to the questions above about pluralistic and contingent analyses. If the two kinds of causation are different, then why are they called the same thing and what to they have in common? On the other hand, if we can clearly set out what it is these two kinds of causation

have in common, don't we have our necessarily true and univocal analysis? The Canberra Plan could offer a solution to this. The conceptual part of the plan gives us the reason why we might call different kinds of causation, and causation in other possible worlds, 'causation' rather than something else. The first part of a Canberra Plan analysis will tell us what they have in common that makes them all causation. The second part of the Canberra Plan will tell us what it is that makes them different kinds, or different realisers, of causation. It will explain what empirically accessible processes causation is identical with, and give us information about causation which may be true, but contingent, or only sometimes true even in this world.

If a contingent analysis of causation is possible, then a pluralistic analysis is just as possible. These approaches come with the same problem, with the same possible solution. Once we are happy with giving an analysis of causation as a role, which can be filled by different actual processes, it doesn't much matter whether the role is filled by different processes in our world, or other possible worlds.

1.7.2 Is causation a relation?

Finally, there is the issue of whether causation should be analysed as a relation or as something else. Many writers take it as read that causation is a relation. The question is what the relata are: are they facts or events, for example? However, there are approaches to analysing causation which do not analyse the relation of causation as such. Both Salmon's and Dowe's accounts of causation analyse the concept of a causal interaction in terms

of a causal process. So first of all we describe what a process is, and then what is different about a causal process. The idea of an interaction is then dependent on the idea of a process.

Cartwright (2004) and Anscombe (1993) both argue that there are certain causal verbs, like pushing and pulling. If causation is to be analysed at all, it will mean finding out what these have in common. We can call this the adjectival approach to analysing causation. There is something – a process, an interaction, a verb – and there are some cases of it which are causal. The business of analysing causation amounts to saying what is different about all these cases of whatever it is which are causal: what do they all have in common which they do not have in common with non-causal cases?

This differs from the relational approach. When we analyse causation as a relation we do not just say that there are relations and some relations are causal. Rather, we think causation is a relation, along with being four feet from and being taller than. When we think of analysing causation as analysing a relation, we are not analysing a special class of relations which are causal, but a specific relation which is *the* causal relation.

If the adjectival approach is the right one, it looks as though it avoids some of the epistemological problems associated with causation. The long-standing puzzle is how we can form beliefs and knowledge about causation without being able to see it or observe it empirically in any systematic way. But if we are not looking for causation as such, but are looking for whatever it is that is causal about certain processes, actions or interactions, then it seems as though the problem was never really a problem. Those processes and interactions are entirely accessible to us empirically, and there's no rea-

son to think that whatever makes them causal also makes them empirically inaccessible to us. By contrast, the relational approach posits the existence of an extra relation besides normal spatiotemporal relations. It implies that something else is going, besides everything that we can see such as contiguity. We then have to say why we think causation, this extra thing, is going on, besides all the empirically accessible events and relations.

On the face of it, this seems like a point in favour of the adjectival approach to causation. It seems to avoid a problem by showing that it was never a problem in the first place, and show that in fact it was just an artefact of how we posed the the question. But it is possible that the adjectival approach does not really remove the problems readily as it appears to. Certainly we can see the causal process, in just the same way that we can see other processes. But it does not follow that we can see what is causal about it. For example, suppose we can see some causal process like a tennis ball flying through the air, hitting the tarmac and bouncing. How do we see that it is a causal process? Under the adjectival approach, the problem can just be rephrased: rather than asking 'how do we see causation?' we can ask 'how do we see that some process is causal?' Some analysis is still required to answer the question.

1.8   Conclusion: empiricism as a motivation for reductive analysis

Reductive analysis can be motivated in two ways. One is metaphysical, to do with having fewer theoretical components in our theory and having the shortest possible list of what there is and ways that it can be structured. But reductive analysis is also an epistemological issue. A concept may be dubious to us because it does not meet our criteria for meaningfulness, or we cannot see a way in which we can know about it. In this case, being unable to observe causation leads us to doubt it on empirical grounds. A reductive analysis can show how causation reduces to something we can observe, and so may be empirically acceptable after all. So here the reductive analysis is epistemologically motivated. I will argue for this in more detail in chapter nine below.

Chapter 2

Relata

2.1 Introduction

Before I begin my survey of different approaches to analysing causation, I want to introduce two puzzles which any account of causation must deal with: the question of what the causal relata are, and what to make of causation by and of absences. In this chapter I will discuss the causal relata question. I will be arguing that our best option is to take the causal relata to be facts.

I want to give a survey of the different answers to the question: "what are the causal relata?" This question is mainly orthogonal to what analysis of causation we are arguing for, and even to what kind of analysis we aim for. For example, a counterfactual analysis such as that of Lewis (2004a) (see chapter four) or Hall (2004) (see chapter 10) may take the causal relata to be events, but so may a physical account such as that due to Dowe (2000) (see chapter 5). Nevertheless, our views on what the causal relata are will affect the way we set out our analysis, and also how we deal with various

41

problem cases. In particular, it will affect how we cope with causation by and of absences, so I will cover this in detail in the next chapter.

There are some assumptions behind the question 'what are the causal relata?' While these are commonly held assumptions there are also dissenting views.

1. Causation is a relation. Mellor (1995, p. 12) denies this and argues that it is an operator. Lewis, in his later work also denies that causation is a relation.

2. The causal relata are all the same kind of thing. Some causal pluralists such as Cartwright might well deny this. Perhaps there could be some cases of causation where the relation is between aspects, and others where it is between events, for example. Another way for this assumption to be false is if the cause is one kind of thing and effect another. For example, in the case of agent causation, the cause is an agent but the effect is an event. (Unlike causation between events and facts where causation is between two things of the same kind.) This is a less common view; although causation is believed to be asymmetric, this asymmetry is not normally located in the cause and effect being different kinds of things.

3. There are two causal relata, the cause and the effect. Several philosophers argue that there are more. Ducasse (1969) argues that causation is a three place relation, with the third relatum being the background

conditions. Philosophers advocating a contrastive account argue that there are four causal relata, with the others being the contrasts to cause and effect (Schaffer 2005). I will discuss these views in more detail in sections below.

While each of these assumptions is denied by somebody, most philosophers accept them, and so I will start with the more mainstream views on causal relata which are based on these assumptions.

First of all, though, I want to discuss why it matters what the causal relata are. One reason is that there are certain features which we might want causation to have, and in order for it to plausibly have those features the causal relata have to have them too. Some cases of this:

- Mind-independence. We probably want a theory of causation according to which it is mind-independent. We want to be able to the say that when rock a strikes rock b, sometime before the evolution of human beings, some kind of causation went on, despite the lack of minds. It is difficult to see how the causation could be mind-independent if it is between causal relata which are mind-dependent.

- Concreteness. If we believe causation is a physical relation, linked to pushing and pulling, we will need to see the causal relata as something concrete and spatio-temporally located.

- Epistemic access. We generally assume that we do have some causal knowledge. This means that the causal relation must be between things

which we can in principle know something about. In general, we can only know as much about causation as we know about the causal relata.

Another reason why we should care about what the causal relata are has to do with our intuitions and the way in which we make causal claims. We want all and only the right cases to count as causation. So any case which is intuitively a case of causation needs to be a case of a relation between the right kind of relata. Or, we need to be able to plausibly rephrase it in those terms. For example, the phrase 'the stone broke the window' could easily be taken as a sloppy paraphrase for 'the stone's being thrown at the window caused the window to break'. So even if at first pass it seems like causation between objects, it is easy to see how we could rephrase it as causation between events. This is the usual constraint placed on most philosophical theories: if the theory yields very counter-intuitive results, we usually think there is something wrong with it.

So there are two criteria here:

- Covers the right cases. Whatever the causal relata are in our theory of causation, we have to be able to account for all the right cases as causal. This is not so hard. Most causal claims can be rephrased in terms of a variety of different causal relata.

- Does not call for implausible paraphrasing. Needless to say, this condition is unsatisfyingly vague. However, there are some clear cases where we would have to paraphrase too much: for example, if we took it that every time we say an event causes another event what we really mean is that God thinking about the first event causes God to think about

the second event, we could safely say that we were paraphrasing implausibly. Other paraphrases, such as from 'Suzie's throwing the rock caused the window to smash' to 'the fact that Suzie threw the rock caused the window to smash' are as innocuous as they come. This is useful condition, but will make trouble for us when we get near the fuzzy plausible/implausible line.

I will argue that no one choice of causal relata meets all these criteria perfectly, and there will have to be a trade-off between some desiderata and others. I will argue that the best candidate for the causal relata are facts: these allow us to be specific enough to cope with problem cases, and allow us to deal with causation by absences. Whilst they seem less actual and concrete than events, if understood as states of affairs they are scarcely more abstract.

2.2 Events

The most popular view is that the causal relata are events (Paul and Hall 2013, p. 7), (Emmet 1984, p. 7). Taking the causal relata to be events is the mainstream view, in the sense that philosophers who are discussing something else about causation, but don't have a particular axe to grind with regard to what the causal relata are, will usually be happy enough to assume that the causal relata are events. So I will begin by discussing events, firstly to explain why they are such a popular candidate, and secondly to explain why I do not think they are the best candidate to be the causal relata. First of all, let's look at the reasons why events are popular.

One big advantage of the view that the causal relata are events is that for any given case of causation we can use the relata to explain why the effect happened when it did. The effect happened because the cause happened right before it. Objects persist over time, usually for longer than events, and so we would have trouble saying why it brought about a certain effect when it did. Certainly we can say that the explosion was caused by the caesium and water, but since the caesium and water were both present before the explosion, that way of phrasing it fails to explain why the explosion didn't happen sooner. If we say that the event of the caesium being dropped into water caused the explosion, then we immediately have an explanation for why the explosion happened when it did.

Meanwhile, events appeal to our desire to have a lightweight metaphysics. Metaphysically committing ourselves to certain kinds of entities can be undesirable. We are metaphysically committed to the existence of whatever we take the causal relata to be, unless we are deflationists about causation itself. It is as if there is a scale of entities which can be posited, ranging from the most concrete and physical to the most abstract and 'spooky'. On that scale, events are fairly near the concrete and physical end. They are spatiotemporally located and extended. They are not mind dependent and they do not sound like anything mind dependent ('facts' still sounds mind dependent, although later I will argue that taking facts to be states of affairs means we can count them as mind-independent). It is easy to think of examples of concrete events which are physically unmysterious, such as rocks hitting windows, or particles of gas bouncing off each other, or a change in an electric current causing a change in a magnetic field. While other events, such as someone

blushing or a stock market crash are less physically straightforward, it is nevertheless easy enough to imagine how they could be made up of smaller events which could be explained in the language of physics. Events seem like they are made of physical parts, such as smaller events, not metaphysical parts, such as a fact being made up of a subject and a predicate. In other words, events are non-spooky. This is a big point in their favour.

However, if we do decide to take the causal relata to be events, we still need to decide what we take events to be. Different conceptions of events will have different consequences for theories of causation. There are several competing views on what events are. The challenge is finding one which does not cause problems in some case of causation. I will briefly look at the options for what events are and discuss what issues each case presents.

One popular view of events, put forward by Davidson (1980) is that they are regions of spacetime. They are defined by time and place but not by properties, so if two apparently different events happen in the same place at the same time those events are identical. We may choose to describe them using different properties but these are just different descriptions: there is only one event in a given place at a given time.

This makes events appealing candidates for being the causal relata, because regions of space time are already in our ontology. They are required by fundamental physics, so we are not adding anything to our ontology for the sake of our metaphysics of causation. This view of events is known as the coarse grained view.

We might feel that while it is attractive to conceive of the causal relata as coarse-grained events which are the same as spatiotemporal regions, this

is also missing features that our causal relata need to have. One thought is that it is the properties of events which fix their effects. Take the case of Suzie throwing a rock at the window, causing it to shatter. The manner in which Suzie throws the rock determines whether the window shatters or not. Throwing it hard might cause it to shatter, while throwing it gently might not. So we might question whether the coarse grained are up to the job of functioning as causal relata.

A different view of events is the fine grained view. According to this view, events are distinguished by their properties, so that if an event happens in a different way, it is a different event. The advantage of this view is that it is more promising when dealing with cases of pre-emption.

Pre-emption cases are a problem for counterfactual theories of causation, and I will discuss them in more detail in chapter 4 where I discuss Lewis's view. However, the short version is that under a counterfactual theory of causation we want to say that causation largely consists in the fact that if the cause hadn't happened, the effect wouldn't have happened. However, in some cases, something 'pre-empts' the cause, so that even if the cause hadn't brought about the effect, something else would have. So in these cases, we think that something is a cause, when there is no counterfactual dependence of the effect on that cause.

A fragile view of events is more promising when it comes to cases of pre-emption. The fragility of an event is how easy it would be for that event to have not happened. For example, if Suzie had thrown her rock at the next window along, would that have been the same throw? What if she had thrown a brick instead of a rock, or a different rock, or thrown the rock at

a slightly different time? On the fragile view, a little change makes for a different event. On a less fragile view of events, it takes a bigger change to make the event a different event.

The window's shattering does not depend on Suzie's throwing her rock, since Billy's rock would have broken the window if Suzie's hadn't. But we can say that the window's breaking in a certain way at a certain time is counterfactually dependent on Suzie's throw. So our answer in cases of pre-emption will depend on what we think events are, in particular how specifically we individuate events.

The fine-grained view is put forward by Kim (1973). It is the view that an event is an individual, a property exemplified by that individual, and a time at which that property is exemplified. On this view, Suzie's throwing the rock, and Suzie's throwing the rock hard, are two separate events, even if they occur at the same time with the same rock.

Another case this makes a difference to is one which brings out problems with transitivity. The example goes like this. Suppose a philosopher has a skiing accident and injures her right hand. She then goes on to write a paper, but she writes it with her left hand because her right hand is injured. The paper gets published. The thought is that the cause of the paper's getting published is the event of her writing it, which she did with her left hand. The event of her writing it with her left hand is caused by the skiing accident. If causation is transitive then if c causes d and d causes e then c causes e. By this logic, the skiing accident causes the paper to be published. But this clearly sounds wrong. Looked at one way this example could be taken to show that causation is not transitive. But another interpretation is

that there is something wrong with the way events are being individuated: we should distinguish the event of the philosopher writing the paper, from the event of her writing it with her left hand. (This example is due toPaul (2000).)

Kim's fine-grained view can help with this, since it makes the event of the paper's being written different from the event of the paper's being written with the left hand. This approach has disadvantages as well, however. It is not ontologically parsimonious: it implies that there are very many more events in the world than we might have thought. It also means there is redundant causation; events are more overdetermined than we might have thought as well. For example, if my teeth are chattering it could be caused by the weather's being cold, the weather's being very cold, the weather's being unseasonably cold, the temperature being below zero, all at once. We would want to distinguish these events if for example Suzie feels the cold less than I do, and her teeth only chatter when the temperature is below zero, but not when it is merely cold. But the cost of seeing events as more fine-grained is that in the case where we do not need this fine-grainedness, the events end up being overdetermined. Once we have distinguished the weather's being cold and the temperature's being below zero as separate events, we have to say that my teeth chattering is caused by both of those events. But this is overdetermination, since either would have done. The root of the problem is that in some cases an event which has been individuated pretty coarsely can sound like a cause, but in other cases we have to split the event into smaller more fine-grained events to pick out the cause.

Davidson (1980, pp. 163-80) had another, earlier view regarding the in-

dividuation of events: rather than distinguish them by their spatio-temporal locations, as with the coarse grained view, or by their location and properties as with the fine-grained view, we could individuate events by their causes and effects. On this view what makes one event different from another is its causal properties.

In some ways, this seems to automatically solve the problems above. For example, in the case of pre-emption we can say that Billy's throw would not have caused the window to shatter if Suzie hadn't thrown her rock. The shattering caused by Billy's throw would have been a different event, because it was differently caused. Menzies (1996, p. 90) argued that if we distinguish events by their causes and effects then the pre-emption problem disappears.

In the case of the skiing accident we can say that the writing of the paper with a left hand is a separate event from writing to paper if we decide that they have different effects. It seems pretty clear that the publishing of the paper is not caused by its being written left handed, so we can say that is a separate event. But there is no need to multiply events unnecessarily: unless the events do differ in cause and effect they can count as the same; their other properties will not make them different.

The worry about this is circularity. It seems distinctly circular to say that the causal relata are events and that events should be individuated by their causes and effect. For this to work we would need our definition of causation prior to our definition of events. It is easy to see how this could lead to circularity of definition. Part of our definition of causation might be that it is between events. Then part of our definition of an event is that it something which is individuated by its causes and effects.

Menzies (1996) argues that there are two readings here: one plausible and one implausible. One reading is that the criterion intraworld: it is a claim about events in the actual world saying that no two events have the same causes and effects. The other reading is that it is an interworld criterion for distinguishing events from other possible events; it is saying that an event cannot have one set of causes and effects in one world and another set of causes and effects in another world.

Menzies (1996) considers the first reading far more plausible. However, he argues, it is not strong enough to do the work we want it to do here. On the weak reading, even if each event is unique in its causes and effects within the actual world, it could still have had different causes and effects. This means that the window's shattering could have been caused by either Billy's or Suzie's throw, so we are back to pre-emption problem.

However, Menzies considers the strong version of this claim to be implausible. It would mean that no event could possibly have had different causes and effects. This makes causal relations necessary. If the window's smashing could only have been caused by Suzie's throw, and Suzie's throw could only cause the window the smash, then the window must be smashed by Suzie's throw. Any world containing one of these events must also contain the other one, and the causal story must play out in the same way in each world. As far as causation is concerned, a world cannot be different unless it has a completely unconnected set of events in it. Since there is, Menzies writes, a widely held view that causal relations are contingent, this is too costly a response to the pre-emption problem.

So there is some debate over what events actually are, and each account

of events holds its own problems for causation: worries about overdetermination, lack of parsimony and transitivity all pull us in different directions on this issue. So although taking events to be causal relata is the norm, when we get into the details it becomes problematic.

2.3 Agents

If we are unhappy with events as causal relata, there are many alternative views. One more unusual candidate for causal relata is agents. This is associated with arguments in favour of free will (Pereboom 2004). The thought is that regardless of determinism or indeterminism, what is really needed for a person to be free is for them to be the cause of their own actions. For that, agents need to be causal relata.

Pereboom describes what he calls 'the problem of the disappearing agent' which agents as causal relata is meant to solve. If factors in the agent's mind bring about her action, whether deterministically or indeterministically, then there is no room for the agent to make a choice between competing motivations. O'Connor (1996) argues for agent-causation on the same basis; it helps us make a convincing case for free will. He argues that we will need a non-reductionist view of causation, and to think of causation in terms of objects having powers or dispositions to bring about certain effects. Agents are objects which bear the dispositional property of being able to make choices.

I have several worries about this view. One concern is that neither writer seems to feel that agent causation is the only kind of causation. (It does seem obvious that something besides an agent can be a cause; when a rock

strikes another rock in space there is not an agent in sight.) So that means that there are two types of causation, agent-action, and event-event. The question we would need to answer would be 'what do these two types of causation have in common which makes them both causation?' and, bearing in mind the motivation for this whole line of thought, 'is that thing they have in common enough for us to claim that agents cause their actions, in whatever sense of 'cause' that's needed to ensure free will?'

Steward (2012) addresses some of these worries in her book 'A Metaphysics for Freedom' in various ways. One way is to say that the agent does not cause the action; the action is the causation. She writes that 'cause' seems a lot like an umbrella term for a set of causal verbs, 'push', 'pull' and so on. (Here she is following Austin; this view is also reminiscent of Anscombe and Cartwright.) If we look at ordinary language, she argues, we can see that it is very natural to talk of substance, particularly agents, as causes. Paraphrasing everything in terms of events to make for some event based metaphysics is, to her, implausible paraphrasing.

Steward points out that there are no cases of agent causation which do not also involve events. It is just that one of the events is an action carried out by an agent. Just mentioning the events wouldn't capture the whole story, she argues, but at the same time cases of agent causation are also causation by events and thus are not mysterious and different.

Steward argues that the real problem with agent causation is that it is a variety of substance causation: causation by substances and not events. The other issue she points out is that agent causation requires the idea that causation takes place on the macroscopic level of agent, rather than the

microscopic level of the particles which make them up, and event between those particles. (These issues of what level causation takes place on will reappear when we discuss Woodward's account of causation, an account which is particularly geared towards causal explanation.)

A big worry for agent causation is the idea that it means two kinds of causation: event causation and agent causation. Steward's reply is that firstly, agent causation is a subspecies of substance causation, and that that is actually quite common. The rock breaks the window. Of course we can rephrase this as the event of the rock striking the window causes the event of window's shattering. But that is true of cases of agent causation too. Steward asks why we should reduce any of these cases to events. We might argue that the rock would not have broken the window if it had not been thrown at it, so the rock didn't break the window; rather, the throw did. But this is strange reasoning: we could equally argue that the event of the rock's being thrown at the window, wouldn't have broken it if the throw hadn't been hard, so the throw didn't break the window, the hardness of the throw did.

Steward argues that what is going on here is that there is usually a trigger event which allows a substance to cause something. Another way of putting that is that there is usually an answer to the question 'why now?' But, she argues, the answer to 'why now?' is not necessarily the cause. It is useful in explanations, but that doesn't mean that it is itself the cause.

In fact, Steward (2012, p. 201) argues that causation is not always a relation, but more of a category, and the causal relata are not always the same kind of thing. Steward has a pluralist attitude to what the causal

relata are.

One concern about agent causation is the 'why now?' problem. This is the worry that if an object such as agent persists for a length of time with the capacity to cause something, then there is explanation for why it brings about the effect at the time when it does. Steward's suggestion of an event which acts alongside the agent as a trigger may be a solution to this problem.

Another concern about agent causation is that it looks like it will create problems for transitivity. If causes and effects are different types of thing, then it seems difficult for one thing to be both a cause and effect, something which is needed for transitivity. Even if there are some cases of simple event causation where transitivity functions normally, it seems odd to say that we cannot have a chain of causes and effects involving an agent.

2.4   Contrasts

So far these approaches have all involved the idea that there are two relata, cause and effect. Some philosophers, however, take a contrastive approach, where causal relation is between four relata: the cause, the contrast to the cause, the effect and the contrast to the effect. For example, Suzie's throwing the rock rather than throwing the feather caused the window to shatter rather than stay intact.

The contrastive approach, suggested by Schaffer (2005), differs from the facts-as-relata approach which I discuss below. When we are using facts as causal relata there are still two causal relata. Those facts might be negative, or they might be contrastive, but nevertheless each fact is one relatum. Ac-

cording to the contrastive theory, however, the causal relata are still events, but causation is a four place relation between four events, two of which are contrasts. One advantage over the facts-as-relata approach is that the relata are still concrete events, not the slightly more elusive facts. However, we do have to note that two of the events involved are still not actual events, but merely possible events. At this point it is a trade off between actual facts, some of which might be negative facts, and concrete events, only some of which are actual. I will discuss the notion of negative facts in more detail in the next section.

An advantage of the contrastive account is meant to be that the contrast for an event is specified within the causal claim, rather than being determined by which possible world is closer. Here's what this means: if we have a non-contrastive account we will make claims like 'Suzie's throwing the rock caused the window to break'. Perhaps we want to know if it is true that if Suzie hadn't thrown the rock, the window would not have smashed (maybe the truth of the causal claim depends on the truth of that counterfactual). That depends on what happened in the nearest possible world where Suzie did not throw the rock. If that world is a world where Suzie threw a brick instead of a rock, then obviously the above counterfactual is not true. On the other hand if the nearest possible world is one where Suzie didn't throw anything, then it is true that if Suzie hadn't thrown the rock the window would not have broken. (Assuming that there was no overdetermination.) So in order to determine the truth of the counterfactual claim we have to find out which possible Suzie-not-throwing-rock world is closest. This is an issue for counterfactual accounts; there doesn't seem to be a really objective

way of determining which world is closest.

The contrastive account is meant to help us here, because the contrast is specified for us. We don't have to find out which other possible world is closest in order to find out what the contrast for Suzie's throwing the rock is. And therefore, we don't need to know which possible world is closest in order to determine the truth of the counterfactual. By making our causal claim more specific, we reduce the amount of information we need about the world in order to find out if this is true. To Schaffer, the contrast is specified by the conversation context, and is part of the semantic content of the claim. If we interpret causal claims correctly, we will see that they are really claims about contrasts causing, not about events causing.

This is related to the issue of how much information is encoded in semantic content of the causal claim as opposed to how much communicated pragmatically. This takes a little unpacking. Suppose we make the following claim:

It is raining.

Normally, we assume that it means it is raining here, not for example on the other side of the world. But there are two schools of thought about how we know that. One approach is to say the the implicit 'here' is encoded in the truth-conditional content of the claim: that is, if we say 'it is raining' but it isn't raining here, we have said something false. Schaffer (2005) took this view. Another school of thought is that the claim by itself does not include anything about where it is raining, and we work this out pragmatically from the conversational context. To see why this is relevant to the questions at

hand about causal relata, let us look at the following causal claim:

> If Suzy had not thrown the rock at the window it would not have broken.

We want to decide if it is true. We know that Suzy did throw the rock and the window did break. Suppose we have some kind of counterfactual theory of causation. Then we need to know what would have happened if Suzy hadn't thrown the rock. The question is whether that contrast is part of the semantic content of the claim. It seems natural to interpret the statement so that the contrast is that Suzy didn't throw anything. But the question is whether this is contained in the semantic content of the claim: so that the claim should really read 'If Suzy's had not thrown the rock at the window, rather than throwing anything, the window would have broken'. Or, whether the claim itself doesn't contain any information about the contrast, and that has to be worked out pragmatically from the conversational context.

This makes quite a difference to the status of the claim. If the contrast is part of the semantic content of the claim, then the counterfactual looks pretty likely to be true. But if it is a matter of pragmatic context then the contrast could just as easily be Suzy's throwing a brick, in which case the window would presumably have broken anyway. If other words, if picking the contrast is a matter of context, then we can't tell if the statement is true or false. If we are basing our account of causation on counterfactual claims, then this is important to our understanding of causation.

Rephrasing the statement in terms of facts about contrasts allows us to ensure that the contrast is included in the semantic content, and thus be

more sure that the counterfactual is true. We can say 'the fact that Suzy threw a rock and not a feather, and that she threw it with enough force to shatter glass, caused the fact that the window broke'. This helps explain why the contrastive approach is more appealing to some than others.

Steglich-Petersen (2012) argues that the contrastive account does not actually help us to determine the truth of the counterfactual 'if Suzie had not thrown the rock the window would not have smashed'. His worry is that we still do not know which is the closest possible world where Suzie does not throw the rock. Our contrast may be 'Suzie walked away rather than throwing the rock' but that could have happened in several possible worlds, and we still don't know which one is closest. Since the question is whether the window would have smashed if Suzie had thrown a feather instead of a rock, we need to know what would have happened if Suzie had thrown a feather. If there are several feather throwing worlds, if we want to know what would have happened, we have to look in the closest one. And for that, we still need an ordering of possible worlds according to which one is closest to the actual one.

I don't think this need be a big worry for the contrastive account, however, for the following reason. When we think of the other possible worlds in which Suzie throws the feather, one of two things could be true. Either the worlds do not differ in terms of whether the window smashes or not, or they do. If they do not differ in terms of which one causes the window to smash, then it doesn't matter which of the worlds in which Suzie throws a feather is closest. If the window doesn't smash in any of them, we don't need to know which possible world is closest, because we know that, whichever possible world is

closest, the window doesn't smash in that world.

If the possible worlds in which the contrast happens do differ in terms of whether the window smashes or not, then our contrast is too broad for us to know if the counterfactual is true or not. Consider if our contrast was something very broad: for example, 'if Suzie had done something else instead of throwing the rock, the window would not have smashed'. In this scenario there are many possible worlds in which Suzie did something else. In some of them the window probably does smash and in others it does not. So, from a contrastive point of view, we can say 'we don't know if that counterfactual is true or not; the antecedent is too broad'. That seems like an acceptable thing to say when the contrast is very broad. There is no reason why it is not also an acceptable thing to say when the contrast is narrower, but not narrow enough. Perhaps the counterfactual claim is 'if Suzie had thrown a rubber ball instead of a rock, the window would not have smashed'. In some of the worlds where Suzie throws a rubber ball, the window does smash, and in others it doesn't; it depends on how hard she throws it. So we should say 'we don't know if that counterfactual is true or not; we need more information about whether we're talking about the rubber ball worlds where she throws hard, or the rubber ball worlds where she throws gently'.

In other words, the contrastive account does not presuppose a ranking of possible worlds according to how similar they are to the actual world. All we need to know is, of the worlds in which the contrast happens, does the window shatter in those worlds? If it shatters in all of them then the counterfactual is false. If it doesn't shatter in any of them then the counterfactual is true. If it shatters in some of them but not others then we need more information

in order to decide if the counterfactual is true or false.

My issue with the contrastive approach is different. A key advantage of choosing events as the causal relata is that events are concrete, actual and spatiotemporally located. In the contrastive approach there are four events involved in any case of causation but two of them are non-actual events. For example, if Suzie throws a rock at the window and breaks it, then the causal relation is really between the cause (the rock throw), the cause-contrast (not throwing anything), the effect (the window's smashing) and the contrast-effect (the window's staying intact). Two of those events do not happen: the not throwing anything and window's not smashing. So causation is then a four-place relation, relating two things in the actual world and two things in possible worlds. This takes away the concreteness and actuality that are meant to be benefits of using events as causal relata.

2.5   Facts or states of affairs

The term 'fact' is ambiguous in philosophy. It can either refer to a true proposition, or to a state of affairs. These make for very different causal relata candidates. If the causal relata are true propositions then taking facts to be causal relata means that causation is a relation between propositions. This makes causation mind or language dependent, since it is a relation between mind or language dependent things. This makes the notion of the causal relata being true propositions unattractive

So, from here on in I will take 'fact' and 'state of affairs' to be synonymous. They are concrete, and either obtain or not. One way to think of facts is

as a combination of an object and a property, related by instatiation. For example, the fact that the window was broken would be made up of the window, the property of being broken and the instantiation relation.

It is easy to see how we can rephrase cases of causation between events as causation between facts. For example, Suzie's throwing the rock at the window causes the window's breaking can just as easily be phrased as the fact that Suzie threw a rock at the window caused the fact that the window broke. In some ways the second way of putting it is more natural anyway.

Say the window smashed because Billy threw a rock at it. 'What caused the window to smash?' could come with several answers which all sound reasonable. 'Billy's throwing a rock' seems like a right answer, and is an event. But 'Billy' could also be a plausible answer; as an agent, Billy could be the cause of the smashing. Another way of answering would be to say 'the fact that Billy threw the rock'. All these sound just as plausible as answers. A very natural way to phrase things is to say that 'Billy's throwing the rock caused the window to smash'. We could interpret this sentence as saying that a relation obtained between events. But the sentence makes just as much sense if we interpret 'Billy's throwing the rock' as 'the fact that Billy threw the rock' and 'the window to smash' as 'the fact that the window smashed'. So, from the point of view of ordinary sentences, there is no reason not to think of the relata as facts. The question is whether facts or events are better candidates for relata on theoretical grounds. Mellor (2004, pp. 309-323) argued that facts are a better choice, although Mellor's view of facts is that they are true propositions.

Armstrong (1997, p. 438) argues that a state of affairs consists in a par-

ticular along with properties which are instantiated by it. He writes that there is a case for states of affairs as causal relata. While the events do the causing, they do so in virtue of their properties.

Facts may have the edge over events as causal relata because of the way they allow us to talk about absence causation, which I will discuss in more detail in the next chapter. If the causal relata are events then when we have a case of not watering the plants causing them to die, we have a case of an absence of an event causing another event. This is problematic for several reasons, chiefly that an absence of an event is not an actual existent thing, and therefore it is hard to see how it could cause an effect. Phrased in terms of facts, however, this causes less of a problem. There is a fact of the plants not having been watered, even though there is no event of the plants not being watered. Negative facts may not be as non-spooky as events, but they are less spooky than absences of events, which is an improvement as long as we want to talk of causation involving absences (and not doing so would be a radical departure from common sense).

2.6   Conclusion: a trade-off

If we look at the different possible causal relata along with their pros and cons we can see that a pattern emerges. Some suggestions are better at satisfying our preference for the causal relata to be mind independent, concrete and physical. At this end of the spectrum we have events. On the other hand, we have kinds of relata which are good for satisfying our intuitions about absences and dealing with cases which would otherwise challenge transitivity.

Facts are good in this way: they make problems with absences less troubling (I will discuss negative facts in the next chapter), as well as allowing us to be specific enough to cope with cases of pre-emption. The trouble is that however much we assert that facts are states of affairs and not something mind or language dependent like statements, facts or states of affairs still seem more abstract than specific events. It still implies an event or object plus some property for it to instantiate, and the instantiation relation is not exactly concrete. So what we gain in power to deal with counterexamples, we lose in concreteness.

That said, facts can offer us the best of both worlds. Facts are made up of a concrete, actual particular and an actual property instance. The instantiation relation is not in itself concrete, but it does not seem like that much of an ideological cost, since it is needed for any particular to instantiate a property and is certainly not something we have introduced just for the sake of causation. Facts allow us to be specific enough to allow us to deal with potential counter-examples to transitivity. In the case of the skiing accident, the fact that the philosopher injures her right hand causes the paper to be written with the left hand, but the fact that the paper is written at all causes it to be published. (Paul 2000).

Below in chapter 11 I will be arguing that there are two kinds of causation: causal influence and causal difference making. The relata for causal difference making are facts. Facts make a difference to other facts. At this level, it is important to accommodate causation by and of absences, and facts do this, via negative facts. The best candidate for causal influence is Dowe's account of causation (I will describe this below in chapter 5). At this level, there

is no causation by absences. Dowe's account is not focused around causal relata; if anything, in Dowe's account the causal relata are objects. I will argue below in chapter 11 that both kinds of causation can coexist and that positing two kinds of causation in this way explains a lot.

Chapter 3

Absences

3.1 Introduction

In this chapter I will discuss causation by and of absences. Like the question
of causal relata, this question is something which is not tied to any one
account of causation, but is rather a general problem with causation. The
problem, put simply, is that it seems very obvious that there are ocases
where an absence causes something, or is caused by something. However,
accommodating these cases in an account of causation proves difficult. At
times I will refer to 'absence causation': this is short for 'causation by or of
absences'.

I will first discuss in detail exactly why absence causation is a problem
for us. I will then go on to discuss the problem of absence causation and
multiplying causes; the worry that if absences can be causes and effects, that
will have the result that there are far too many. I will then discuss ways in
which a philosophical commitment to the causal relata being one thing or

another could affect what we end up saying about absence causation. Next I will discuss the possibility that it is thinking of causation as a relation which creates part of our problem with absence causation. Finally, I will discuss the worry that absence causation forces us to produce an analysis of causation which is disjunctive.

My view is that the best way to approach absence causation is to treat it as causation by or of negative facts. I argue that we should use the understanding of negative facts due to Jago et al. (2012) for this purpose. I will argue that worries about absences specific to causation's being a relation are misplaced, as are worries about disjunctive analyses of causation.

On the view set out by Jago et al. (2012, p. 121), facts can be located, including negative facts. Jago and Barker argue that a negative relational fact is located where its relata are (just as with a positive relational fact). So the fact that I did not water the plants is located where I am, and where the plants are. I do not think we should be concerned that this appears to be in two discontinuous spatiotemporal locations, after all so do some positive particulars, such as a theatre performance with an interval. Nor do we need to worry if only part of the fact is spatiotemporally located appropriately relative to the other causal relatum (ie the effect should be in the cause's forward lightcone (Dowe 2009)). Again, I can take the example of a play. I can say during the interval, truthfully, that a play moved me, even though some of the play is still in the future. Perhaps it would be more precise to say that the first half of the play moved me. But the original statement is not false. And there is nothing physically perplexing about it, as long as at least part of the cause is temporally prior to the effect.

3.2 The problem with absence causation

There are several issues surrounding absence causation, some which arise when it is combined with other theses about causation such as the nature of the causal relata, and some which are just a result of the very idea that absences can be causes and effects.

Cases of absence causation are ubiquitous. Here are some examples:

1. I don't stir the soup. This causes it to burn.

2. A person is beheaded. The beheading causes an absence of blood flow to their brain. This absences of blood flow causes death.

3. I don't water my plants. This causes them to die.

Not to account for causation by and of absences somehow in our analysis would mean a major departure from our everyday concept of causation. These cases are intuitively cases of causation. To deny that there can be causation by absences is to deny that we can cause death by decapitation; the absence of blood flow to the brain is the direct cause of death in that case (Lewis 2004b).

In this section I want to disambiguate the reasons why causation by and of absences might be a problem for us. This is important since where the worry comes from will determine how we cure or avoid it. For example, if it is bound up with our choice of causal relata, then we will not make any progress with the problem by changing our analysis of causation, and vice versa.

1. General intuition of perplexity. It is hard to see how an absence can cause anything. There is nothing there to do the causing.

2. Problems specific to causal theory. Eg Dowe's analysis of causation means that causation can only happen when a quantity of energy or momentum is conserved, and an absence cannot instantiate a quantity of energy / momentum. (Dowe 2000)

3. The problem of multiplying causes. This is the worry that if absences can be causes, then there are very many more causes than we would have thought. For example, the absence of a meteorite landing on my head is a cause of this thesis.

4. Problems specific to the causal relata. This worry is based on the idea that if the causal relata is x, absence-x either can't exist, or if it does exist, can't have causal efficacy.

5. Problems specific to causation being a relation. That is, the idea that one of the relata's being missing is a problem for thinking of causation as a relation.

6. Problems relating to disjunctivism. This is Lewis's worry: if causation is all about biff, then how can we characterise absence causation in terms of biff without our analysis becoming disjunctive (Lewis 2004b).

7. Problems due to possible causation ("would-cause" counterfactuals). (Dowe 2009)

I will not devote much discussion to (1). It is too vague to argue with. I

suspect that on closer examination this general feeling of 'that can't be right'
will turn out to be one or more of the issues below, and that when these are
addressed the general intuition that absence causation is a problem will go
away.

The second kind of worry, problems specific to a particular analysis of
causation, I will consider in the chapters concerning those analyses, when
it becomes relevant. Most importantly, I will discuss Dowe's approach in
chapter 5 below.

The next three sections will discuss (3), (4), (5) and (6) above in turn.

3.3 Multiplying causes

One problem with absence causation is that it seems to lead to too many
causes. Take the case of Suzie throwing her rock at the window. If absences
can be causes then we may have a hard time saying why the absence of
shutters over the window is not a cause of window's breaking. Not to mention
the absence of a sudden hurricane, or Suzie's having a stroke before she can
throw her rock, or any number of other things.

Another way of putting it is to say that normative considerations seem to
play an unreasonably big part in our judgements about when an absence has
been a cause. Here is how that works: suppose in the example above, that
there are usually shutters over the window, and they just happen to have
been left open today. Then it seems much more reasonable to describe the
absence of shutters over the window as a cause of the window's breaking. But
the absence of a sudden hurricane still seems like a bizarre thing to pick out

as a cause. The difference is that the presence of the shutters was normal, but there is not normally a hurricane. Including considerations of what is normal in our causal judgements seems like a bad idea: besides making causation non-intrinsic, it also makes it subjective, which may be something we would look to avoid.

If we do not want to include normative considerations in our causal judgements, we have to find some other way of ruling the absence of shutters in as a cause, and the absence of hurricane out. Either that, or we have to say that both are causes, and that there are very many more causes than we originally thought; or, alternatively, we have to say that absences cannot be causes. Different accounts of causation come with different approaches to this problem. Strevens (2013) has possible solution to this, as do Handfield et al. (2008), discussed in chapter 11.

3.4 Absence causation and the causal relata

In this section I will look at why the ways in which our choice of causal relata can help with the issue of absence causation. I suggest that this will play a big part in our choice of causal relata: causal relata which cannot accommodate causation by and of absences look likely to be the wrong choice.

3.4.1 Events and absence causation

It is very natural for us to think of causation as an actual physical relation between particular events. But this comes into conflict with the idea that absences can be causes: it is hard to see how an absence can physically affect

something when it is not there. The more concrete and physical our under-standing of causation, the harder it is to make room for causation by and of absences. Understanding the causal relata as events has the benefit that it makes the causal relata something physical and concrete, with a location in time and space and properties. The location means that we can answer questions of the "why here and why now?" sort. The properties mean that we can answer "why this way?" questions. Events are specific; they are particulars.

One problem is individuating absences; answering the question of when to absences are the same and when they are different. The event of my watering the plants the event of Bill watering the plants have different properties, as well as different locations in time and space. But the non-event of my not watering the plants and Bill's not watering the plants do not have their own properties and it's not clear whether they are part of the same absence (no one's watering the plants) or whether they are different. Another worry (due to Mellor (1995, p. 316)) is that the properties of a cause, along with the laws of nature, fix its effects. But with absences having no definite properties, we do not know what effects to attribute to them. For example, the gale-force wind caused the ship to be wrecked, but the steady wind caused the ship to make progress. But if there is no wind, we do not know if an absent gale-force wind that causes the ship not to be wrecked, or an absent steady wind caused the ship to stay still.

It is not clear where an absence is located. For example, if I do not water the plants, where and when is the absent-event of my not watering the plants? Perhaps it is located wherever I am, when I am doing something

other than watering the plants, or perhaps it is located in the same place as the plants. The absence could be thought of as in several different places in time too: perhaps all the time in which I am not watering the plants, or perhaps the time when I did water the plants in the nearest possible world where I do water them. One way to individuate absences might be to specify their spatiotemporal locations. This would allow us to distinguish my not watering the plants on Tuesday and my not watering the plants on Thursday, which might be important when absences are involved in causation ("The plants died because, although you watered them on Thursday, they needed watering on Tuesday.") (Davidson 1980).

Nevertheless, the main advantages of taking the causal relata to be events (that events are concrete and spatiotemporally located particulars) seem to disappear when one or more of the relata is an absence. So with absence causation in mind, events do not look like the best candidate for the causal relata.

3.4.2   The contrastive approach

In the previous chapter I discuss the contrastive approach to causation: that is, each causal claim is a claim that c rather than $C*$ caused e rather than $E*$. Rather than causation being a two place relation between cause and effect, causation is a four place relation between cause and its contrast event and the effect and its contrast event.

Schaffer (2005) argues that contrastivity can solve the problem of absences and causation. He sums up the problem like this. The main problem is

the question of what absences are. Are they (a) nonactual events? (b) actual facts? or (c) actual events, in the sense of the events that happened instead? To give examples, (a) is the event of my stirring the soup, which doesn't occur, (b) is the fact that I don't stir the soup and (c) is the actual event of my answering the door instead of stirring the soup. Schaffer takes (a) and (b) to be metaphysically spooky: nonactual events because they are nonactual and facts because they are abstract. (c) meanwhile conflicts with the counterfactual dependence account of causation, since the soup's burning does not counterfactually depend on my answering the door. It is not necessarily the case that if I had not answered the door, the soup would not have burned: for example, if I was asleep I would not have answered the door, but the soup would still have burned.

Schaffer (2005) argues that a contrastive account of causation can fix this apparent paradox. When we talk of absences we are referring to the actual events which take place instead of the absent event. We talk about the absent event to set the contrast class. So when we say 'my not stirring the soup caused it to burn', we mean that my answering the door caused the soup to burn. But we say 'not stirring the soup' to make it clear that it was specifically my answering the door *rather than* stirring the soup which caused the burning.

I disagree with Schaffer on a number of points. Firstly, I think that facts understood in the right way are not unacceptably abstract. So understanding absences as actual facts is a perfectly legitimate option. So Schaffer's proposed solution is not needed. Secondly, Schaffer's proposed solution does not fix the problem. The question is what the right answer is to 'what caused

the soup to burn?' According to Schaffer's account the answer would be 'my answering the door rather than stirring the soup caused it to burn'. But I don't think that's quite right. My answering the door has nothing to do with the soup burning. Even if we were to take events to be the causal relata, we would have to admit that is the properties of the cause events which set how the effects happen, when they happen and sometimes whether they happen at all. Our answer to why the soup burned should include the event which triggered the burning, and the properties of that event which set how the event of the soup's burning happened, and when. Here is one way of thinking about it: changes in how I stir the soup will effect the soup (eg if I stir it carelessly it might burn anyway) but changes in how I answer the door (quickly, slowly, rudely etc) cannot directly effect the soup. Schaffer's account encourages us to appeal to something actual and concrete at the cost of appealing to the thing which is *relevant*. The negative fact of my not stirring the soup may be 'spooky' but at least it is connected to event it is supposed to be causing.

There are some other worries about the contrastive approach to causation. One issue is that it makes all causation absence causation. For all cases of causation, there will be two relata which are not actual. In a positive causation case, these are the contrasts in a negative causation case they are the cause and effect, but either way, there are absent events involved as relata in the four place relation. Thus if our issue with absences was that they are metaphysically spooky in some way, we have just made all causation metaphysically spooky.

Another issue is that there could be cases where the contrast is also an

absence. Suppose we have a case where I am trying to start a campfire, and can't manage it. There could be a variety of reasons why: lack of heat (if I can't get a spark), lack of the right fuel (perhaps the twigs are damp) or lack of oxygen (fuel so tightly packed that the air can't get where it needs to be). And I might easily find myself saying that the cause of the lack of fire is lack of heat rather than lack of oxygen, for example. In this case both the cause and the contrast are absences. Now Schaffer would respond that this isn't the right contrast: the right contrast would be something like my scraping the match along the strike paper (which would be what I would be doing instead of getting a spark). But the only reason to pick this as the contrast seems to be that we have stipulated that if the cause is an absence, the contrast to the cause must be something positive. There will be many cases where intuitively the contrast to an absence cause is another absence.

3.4.3   Facts and absence causation

In this section, I will discuss the ways in which taking facts to be the causal relata can help us understand cases of absences causation. I will argue that, especially for dealing with absence causation, they are a much better choice than either events in the standard two place relation or events as part of a four place contrastive relation. As in the previous chapter, I am taking facts to be concrete states of affairs, usually consisting of a particular, a property and the instantiation relation. I am not talking about true propositions.

If the causal relata are facts then cases of causation by or of absences will be cases of causation involving negative facts. For example, rather than

there being a causal relation between the event of my not watering the plants and the plants' dying, there is a causal relation between the fact that I did not water the plants and the fact that they died. So instead of one of our relata being an absent event, one of our relata is a negative fact.

Taking the view that the causal relata are facts helps us with the problem that absences are non-specific about their place, time and properties. Event absences are unspecific about time and place, as well as other properties. Negative facts can help us here. Negative facts can include as much information about time and place and the details of what did not happen as we wish.

Mellor (2004) argues that negative facts are less metaphysically troubling than negative particulars (such as event absences) for the same reason. The problem with negative particulars is that they are unspecific. Take the event of my failing to stir the soup. Is that a case of my failing to stir the soup for a long time or a short time? Mellor argues that because there is no answer to this question, event absences are metaphysically suspect. They don't have properties, and Mellor argues there is no adequate way of specifying exactly which absence we are talking about. Facts do not have this problem. We can say that the fact that I didn't stir the soup for a certain length of time was the reason why it burned, or we can say that the fact that I didn't stir the soup at all was the reason, or that I didn't stir it often enough or stirred it at the right time. These negative facts are detailed enough that we can distinguish one from the other, and they have properties which, according to Mellor, makes them less metaphysically spooky than event absences.

But there are also some worries about using facts, including negative facts,

as causal relata. One is that facts are sometimes considered more abstract than particulars. If facts are true propositions then they are things which are made true by the world, but are abstracted from it. However, according to Schaffer (2009) facts are synonymous with states of affairs, and are made up of a particular instantiating one or more properties. On this reading, there is no reason to see facts as more likely to be considered abstract than concrete particulars. The only ingredients of facts are concrete particulars, properties and the instantiation relation, none of which are abstract. [1]

In their paper 'Being positive about negative facts', Jago et al. (2012) argue for this view of facts as states of affairs. Negative facts, according to Jago and Barker, are not negated facts or facts involving negative properties, but particulars anti-instantiating certain properties. So negative facts do not include an absent particular or negative property: rather, the thing which makes them negative is that instead of involving the instantiation relation, they involve the anti-instantiation relation. This has the advantage that each negative fact still involves a positive, non-spooky particular and a positive, non-spooky property. The only strange part is the notion of anti-instantiation, which is hardly more strange than the notion of instantiation itself.

---

[1] Another worry is due to Molnar (2000), who argues that there are no clear truthmakers for negative facts. Though Beall (2000) offers a counterargument to this which he calls the polarity approach. Armstrong (2004) offers another option based on the notion of totality. What makes negative facts true is the whole world *totals* everything that there is, implying that there is nothing else. This view is partly defended by Tallant (2010).

3.5   Absence causation and the relational nature of causation

In this section I will discuss problems with absence causation which arise from causation's being taken to be a relation. Sometimes worries about absence causation are raised in a way which has nothing to do with the nature of causation, or with the specifics of our analysis or what the causal relata are. Sometimes the concern is to do with the claim that causation is a relation. A relation needs to take place between two or more relata. If one of the relata is absent, that sounds like a problem. Lewis had a worry along these lines which I will explain in detail.

3.5.1   Lewis's worry

This is one of the worries raised by Lewis (2004b) in 'Void and Object'. If causation is to be a relation, there must be two relata. He offers an example: suppose I am cast into the void and as a result die. The void is precisely an absence of anything at all. It looks as though it certainly could be a cause: specifically the absence of pressure would cause me to explode and cause my blood to boil, while the absence of air would cause me to asphyxiate and the absence of heat would cause me to freeze. This is a clear case of causation. However, Lewis argues that it cannot be an example of the causal relation. There is nothing (literally) to be the other relatum. Liebesman (2011) dubs this objection 'the Missing Relatum objection'.

Lewis writes:

So I insist, *contra* Menzies, that causation cannot always be the bearing of a causal relation. No theory of the causal relation, neither Menzies's theory nor any other, can be the whole story about causation.

(Lewis 2004b)

Missing relata look like a problem for a relational analysis, since there certainly are cases where causation appears to have an absences as either cause or effect or both. Lewis sets out four options for coping with this type of case:

1. Say that absences don't cause anything; Lewis thinks this is absurd.

2. 'Reify absences non-reductively': this amounts to saying that absences are real. Again, this is an unattractive option for a philosopher committed to saying that everything in the world is physical.

3. Say that holes can be identified with hole-linings, which means the void can be identified with everything that surrounds it. I will talk about this possibility more below.

4. There is no causal relation with the void as relatum; rather causation is not a relation even though we can describe it in terms of relations between events. This appears to be Lewis's view in 'Void and Object' (Lewis 2004b):

Lewis favoured option (4). Lewis goes on to write that an advantage of a counterfactual account of causation is that 'when the relata go missing,

it can do without any causal relation at all.' (Lewis 2004b) It is hard to understand what causation could be, however, if not a relation.

Liebesman argues that we do not need to worry about this, since there are other cases where we seem to have a relation where one of the relata is absent, such if I fall into a hole. In this case, the relation is between me and a hole (an absence in the ground, we can assume), and the relation is one of falling into. One of the relata is an absence, and yet, Liebesman says, there is no metaphysical worry here.

I think that this comparison between absences as they figure in causation and holes deserves more scrutiny. In the following section I will look at various philosophical approaches to holes and see how these might apply to absences in a causal context.

3.5.2   Holes

In his paper 'Causation and the Canberra Plan', Liebesman (2011) argues that absence causation is similar to cases such as falling into a hole, where the relation "falling into" relates two things, one of which is a hole. He argues that absence causation is no more a metaphysical problem than falling into a hole.

This argument by analogy stands or falls how analogous the two things being compared really are. The question here is how much absences and holes are alike in terms of our metaphysical treatment of them. The situation is complicated by the fact that we can see holes as a special case of absences: as absences in something. So as well as being similar holes and absences also

share some of the same extension. The screw hole in a plank of wood is an absence. But the absence of anyone watering the plants is not something we would normally see as a hole. For Liebesman's argument to work, holes and absences have to be alike enough (or co-extensive enough) for us to effectively apply a philosophical treatment of holes to absences.

To find out if there is any mileage in Liebesman's argument, I will be looking at the common theories of holes, and whether they can be applied to absences in case of causation. It will become apparent that none of the metaphysical approaches to holes work for absences in general, at least not for the kinds of absences we are usually talking about in the case of causation. So I will conclude that Liebesman's argument is too speculative. Without some explanation of why falling into a hole is non-problematic, and how this can be applied to causation by absences without its needing to be altered beyond all recognition, Liebesman's argument doesn't offer us any reason to stop worrying about the missing relatum objection.

Theories of holes

There are a variety of ways of dealing with holes. The issue is that we talk about holes as though they are real objects which can move around and be part of relations like 'fall into'. First of all, there are the approaches which rely on holes being holes in something:

- There are no holes, just holed objects.

- Holes are their linings

Two further approaches treat holes as essentially negative:

- Holes are negative parts.

- Holes are 'genuine' absences which are truthmakers for negative facts.

Finally, there are two approaches where we define holes spatiotemporally, in much the same way as we might define events:

- Holes are regions of spacetime.

- Holes are qualified regions of spacetime.

So is there any scope for adapting one of these approaches to holes to the case of absence causation? In the rest of this section I am going to go through these approaches one by one and see how applying them to absence causation might pan out.

Firstly, there is the approach to holes according to which they are genuine absences. Importing this approach to absences does not solve any problems. We still have exactly the same problems we started with, particularly the worry that absences cannot be causally effective when there is nothing to them. Reified absences are metaphysically problematic: for this reason it is not a very appealing way to cope with holes, and it is no better for absences in a causal context.

Next, I will look at the approaches to holes which are based on there being holes in something. Tucholsky (1930) writes: 'There is no such thing as a hole by itself'. It seems like this is exactly what an absence is. However, I think that these approaches shouldn't be ruled out immediately: in the same way that a hole is a hole in something an absence can be an absence in something or of something. One approach of this type is the view that

we should re-interpret claims about holes as claims about something having the property of being holed. Instead of there being a hole in the piece of cheese, the piece of cheese has the property of being holed. To import this to absences, instead of their being an absence of my watering the plants, my watering the plants schedule has the property of being interrupted. Now, in a previous section I have briefly discussed the possibility that the causal relata could be properties. But it would definitely be strange if causation was a relation between properties in the case of absences, but not in the case of normal events. This would mean absence causation was different from normal causation and it would be hard to see how we could preserve transitivity. We want to be able to say that *a* causes *e*, when it does so via a chain of events, *b*, *c* and *d*, which include an absence (Paul and Hall 2013). So the view that there are no holes only holed objects will not help us with absence causation.

The holes as hole linings approach however looks a little more promising at first. Trying to adapt this to the case of causation leads to strange results. If the causal relata are events, we would have to be looking at the 'linings' of absent events. For example, the light's being on before and after the light going on would become cause of my not being able to see when the light is out, since the absence (its being dark) would under this approach be identified with its "lining" (the lights being turned out and then later turned on again).

In the context of causation holes as hole linings translates to absences as absence boundaries. When we think of holes in terms of holes in pieces of cheese, and we start talking of hole linings, we will sound as though we mean

the very thin layer of actual cheese which surrounds the hole. But there is another way of thinking about linings: as the infinitely thin boundary of the absence. So in the case of a hole in a piece of cheese, this would mean the infinitely thin boundary between the cheese and the air filling the hole; it doesn't actually contain any matter, but is just a surface. This makes a lot more sense for absences which are not really absences in anything. For example, suppose Suzy plans to throw a rock at a window but thinks better of it. We could draw a spatiotemporal boundary around Suzy's absence of throw, perhaps from her hand, following the trajectory the rock would have taken to the window, around the width of the rock she would have thrown, and around a given time, perhaps the time during which Suzy considered throwing the rock. There is no 'lining' which is made of anything, but nevertheless we can imagine an infinitely thin boundary which picks out that absence of throw.

Returning to the example of the light's being turned off, say the light is on between $t1$ and $t2$. At $t2$ it is turned off and at $t3$ after that, I can't see. Then later at $t4$ the light is turned on again until $t5$. This means that my not being able to see at $t3$ is caused by the light's being turned off at $t2$ and on at $t4$. This means something at $t4$ is partly the cause of something at $t2$, which sounds unintuitive.

Looking at this example, it seems pretty strange that part of the cause takes place after the effect. However, this is not really so uncommon. For example, if I put a cake in the oven to bake, the cause of the baking is the oven's being hot. The oven carries on being hot for some time after the cake is baked, so in some sense part of the cause comes after the effect. This does

not seem paradoxical. It would only be a problem if all of the cause was after the effect, for example if I put a cake in the oven, it bakes, and only after that do I even turn the oven on.

However, there is still something strange about identifying the cause of my not being able to see at $t3$ with the lights being turned off at $t2$ and on at $t4$. It looks as though that implies action at a temporal distance. After all, there is no chain of events linking the light's being turned off at $t2$ with my not being able to see at $t3$. At least, there is no chain of events which doesn't include its being dark between $t2$ and $t3$, which would be re-introducing the very element we are trying to remove. The idea of identifying absences with their spatiotemporal boundaries does not work.

Here is one hypothetical way someone might try to defend the 'absences as absence linings' approach. Perhaps instead of temporal linings we could think of modal linings. Where there is a gap in an object it makes sense to contrast that gap with the object on either side, naturally giving us the idea of identifying it with the 'lining'. But in the case of an absence of event it is not so clear whether it should be contrasted with the events (or processes?) on either side of it, or with the event which is absent. In the case of a hole in a piece of cheese, the obvious thing to identify the hole with is the cheese surrounding the hole. The obvious contrast is a spatial contrast. But with an absence of event, it not clear whether we should look for a spatio-temporal 'lining' or a modal 'lining'. It seems to depend on the case. If our absence of event is the interval in a concert, for example, it would be most natural to contrast this absence of music with the music being played before and after the interval. But if it is a case of Suzy not catching the rock Billy throws

at the window, it seems most natural to contrast this with another possible scenario in which Suzy did catch the rock. It is not clear what the right contrast class for an event absence is.

We could look for a spatiotemporal contrast or a modal contrast, and both come with problems. This is a general issue with how we define and individuate temporal absences. This worry is not unique to event absences, but is also a problem for spatial absences, holes. Take a screw hole in a piece of wood. The hole lining is made of wood, provided we look for a spatial lining. But it seems to make almost as much sense to look for a modal lining: if there wasn't a hole there, there would be a screw there instead. So the hole should not be contrasted with the wood, but with the screw. We should describe the hole as an absence of screw, not an absence of wood.

Identifying absences with their modal linings comes with its own problems. The contrast is then not actual, and it is not so clear that a non-actual event is better than an actual absence of an event, from a metaphysical point of view. We would have to know what possible worlds are closest in order to know what the lining is. And the causal relation would then sometimes be relation which holds between events which are in different possible worlds, respectively. The idea of modal linings opens a bigger can of worms than the missing relatum objection did in the first place. So the idea of identifying absences with either their spatiotemporal or modal linings seems unpromising.

Similarly, there is the view that holes are negative parts. This is also unpromising. It has all the problems of both the preceding views. Firstly, we can ask 'parts of what?' Absences in the case of causation are not clearly

parts of anything. In the same way as with the hole linings approach we first have to find something for the absence to be of or in, under the negative parts approach we would have to find something for absences to be a part of. Secondly, the idea keeps absences as negative, which gives us the same problems of metaphysical spookiness and wondering how something which is not there can have any causal efficacy.

The remaining approaches to holes suggest that we either view them as spatiotemporal regions or as spatiotemporal regions with properties. One thing worth noting about these approaches is that they have also been taken with regard to events. This offers the possibility of a unified approach, so that we can take the same approach to causation whether or not it is absence causation. By specifying a spatio-temporal region, we get the where and when, and by specifying properties, we get the how. This seems like a promising approach to both holes and absences. However, I suggest that it would need some modifications. Take the example 'I fell into a hole'. We could rephrase this as 'I fell into such-and-such a region, with the property of not being full of earth'. The only issue here is that 'not being full of earth' is a negative property, something to be avoided if possible. One answer might be to replace it with 'being full of air'. But this runs into the problem of relevance: the hole is something I fall into not because of what does fill it, but because of what doesn't. In any case, this solution will not work if the hole contains a vacuum (never mind how the vacuum exists without the surrounding matter filling it; we will just assume for now that this is somehow possible). So the approach of using spatiotemporal regions with properties seems unsatisfactory even in the case of relations where one is a hole. Let

us consider a case of absence causation: the loss of blood pressure to the brain causes death. Suppose we rephrase this as 'the spatiotemporal region of the arteries in my head, in a certain period of time, had the property of containing less than a certain amount of blood' or possible '... had the property of having lower than a certain amount of pressure'. Again, this sounds reasonable, except for the use of a negative property. This way of dealing with holes, then, seems to deal as badly with absences in the causal context as with holes.

My suggestion, again, is to use facts in the sense of states of affairs. This is very similar to the approach above. The facts contain a particular (which can be individuated by picking out a spatiotemporal region) and a property. The difference is that facts also include an instantiation relation. According to Jago et al. (2012), in the case of negative facts this is replaced with an anti-instantion relation. This means we rephrase the above example about low blood pressure like this: 'the spatiotemporal region of the arteries in the head, in a certain period of time, anti-instantiated the property of being full of blood at at least a certain pressure' (where the certain period of time is the time shortly before death, and the certain pressure is whatever pressure is needed to sustain life). This has the advantage of citing the correct, relevant property. It is not what pressure the blood is at which causes death, but the pressure's falling short of what is needed to sustain life. But this also avoids having to refer to negative properties, or event absences. The ingredients of the fact are an actual particular, picked out spatiotemporally, and a positive, actual property. These are decidedly non-spooky ingredients. We do need the instantiation and ant-instantiation relations, but these are just two relations,

which is surely preferable to positing a full palate of negative properties.

3.5.3 Summing up

To conclude this section, one approach to dealing with holes almost works for absences in a causal context. However, I have argued that negative facts are still a more appealing option.

With regard to Lewis's concern about missing relata, my view is that the problem would benefit from being rephrased. It is not the case that there is a relation, and one of the relata goes missing. This would be the case where Suzie throws her rock and the vase and this causes the vase to shatter, but the event of Suzie's throwing the rock disappears. In that case, we would have a problem, not just in explaining what caused the vase to shatter then, but also in saying what the causal relation was meant to hold between. This is not the kind of case we or Lewis are talking about. Instead, we are talking about cases where both relata are there, but one (or more) of them is negative, such as the case of low blood pressure causing death.

With this in mind, the question is how to cope with the negative when it forms one or more of the relata of a relation. It is not necessarily the case, as Liebesman argues, that just because we have other relations with a negative relatum (or two) and we do not worry about them, that we shouldn't worry about causation either. There may be ways in which the causal relation comes with unique problems. We have seen that approaches that plausibly work for holes do not work for causation. On the other hand, the view of holes as spatiotemporal regions with properties nearly works for holes and

absences in a causal context. The approach where we understand absences as negative facts, however, is still more promising.

3.6   Absence causation and disjunctivism

This is another related problem with absence causation which was raised by Lewis (2004b), in his paper 'Void and Object'. It relates to the idea that we might be able to define causation in terms of biff. Biff is a term used to describe a kind of physical causal link in which one thing pushes another. Because of absence causation, we could not define causation in terms of biff by simply that c cause e iff c biffs e. Instead, we would have to have several disjuncts to our analysis: c causes e iff ...

This is specific worry about absence causation, then: about the way in which it affects our ability to define causation in terms of biff. In some ways, this perhaps deserves to be classified under worries about the way in which absence causation interacts with specific theories of causation. However, wanting to analyse causation in terms of biff is more general than any specific theory. It is a very natural and intuitive thing to say: that causation has something to do with push, pull, force, transference of energy, in some way. Ruling it out rules out an entire category of approaches to causation. Hence, the topic has its own section.

Any definition of causation in terms of biff, according to Lewis, would end up being disjunctive. Instead of saying what unifies causation, as Lewis thinks we should with an analysis, we are giving a set of separate criteria, any one of which would be enough to make a case causal.

I agree with Lewis that a disjunctive analysis is a failed analysis. In my view the analysis relation is connected to a grounding relation. When we say that *a* can be analysed in terms of *b*, we are saying that facts concerning a are grounded in facts concerning b. So, a disjunctive analysis looks something like this:

- *a* can be analysed in terms of *b* or *c*. Thus,

- facts about *a* are grounded either in facts about *b* or facts about *c*.

This does not really help us understand the concept *a*, or facts about *a*. If facts about *a* do not really have anything in common in what grounds them, then it does not seem as though there really is a concept *a*. So, a disjunctive account would be a problem and we need some way to reply to Lewis's worry.

Set out in distinct steps, Lewis's argument looks like this:

1. We might want to define causation in terms of biff.

2. We cannot describe an instance of causation by/off an absence in terms of biff *in the same way as* we could describe an absence-free instance of causation in terms of biff.

3. Therefore, if we were to define causation in terms of biff, we would have to give separate sets of conditions, joined by a disjunction.

4. This would result in a disjunctive account of causation, which would be a failed account.

Lewis's approach is to say that we should not try to define causation in terms of biff. However, for reasons I will outline in future chapters, especially

5 and 9, I think that biff or something like it should play a part in our analysis of causation. I agree with (2). In my view the place to push is the move from (2) to (3).

Whether or not (3) is true depends on the details of the account. The toy account Lewis suggests would indeed be unsatisfactory. I think it is possible though that we could have an account that did the job without providing a disjunction. In chapter eleven I will suggest that Strevens's account of causation could be capable of doing this. In any case, Lewis has not shown in 'Void and Object' that no account of causation in terms of biff could possibly be unified and not disjunctive. He has just shown that the most obvious way of approaching it wouldn't work.

3.7   Problems due to possible causation

Dowe (2009) argues that there is a hitherto unnoticed problem with absence causation due to a specific kind of counterfactual: counterfactuals *about* causation, that is, would-cause counterfactuals. For example, "if Suzie threw her rock, it would cause the window to break".

Dowe writes that it is important that we have what he calls A-P independence. That is, logical independence between claims about actual causation and claims about possible causation. He argues that absence causation violates A-P independence, and that doing so brings problems.

Firstly, let's look at why Dowe thought absence causation prevents claims about actual causation from being logically independent from claims about possible causation. Dowe's logic runs something like this:

1. If Suzie had thrown the rock it would have caused the window to break.

2. If it is true that Suzie threw the rock and it caused the window to break, then it follows logically that the window did break.

3. So, if Suzie threw the rock the window would have broken.

4. Suppose we represent 'Suzie threw the rock' as a and 'the window broke' and b.

5. Then if not-not-a then not-not-b. (Since double negatives cancel out.)

6. In counterfactual accounts of causation, we usually say that 'if not-c then not-e' implies 'c caused e.

7. If we accept both this and the previous point, we get 'not-a caused not-b'. Or, 'Suzie's not throwing the rock caused the window to not break'.

Dowe's point is that if we allow causation by absences we cannot avoid this logical link between claims about possible causation and claims about actual causation.

Next, Dowe argues that this logical dependence between possible and actual causal claims is a problem, because it means one of two unpalatable options.

Either, absence causation supervenes on patterns of possible causation. Dowe writes that "therefore absence causation is in some sense a kind of parasitic, second rate causation" Dowe (2009, p. 28).

Or, possible causation is redundant, since any story about possible causation can be replaced by a story concerning actual causation, using the appropriate claims about actual causation. Dowe argues that possible causal claims should be taken more seriously than that. We need to be able to make would-cause claims; they are not just a redundant way of phrasing claims about actual causation.

My response is that we should accept the either. I think we need not consider absence causation as 'parasitic' or 'second-rate', but simply as derivative, in just the same way that many (even most) of the things we care about philosophically are derivative. I will argue in chapter 11 that the kind of causation which allows for causation by absences is different from the kind of less derivative causation which Dowe describes. For now, I think the best response to Dowe's worry here is to accept that absence causation is derivative.

3.8   Conclusion

Let us return to our five reasons to worry about causation by and of absences and see whether these have been addressed.

1. General intuition of perplexity. It is hard to see how an absence can cause anything. There is nothing there to do the causing.

2. The problem of multiplying causes. This is the worry that if absences can be causes, then there are very many more causes than we would have thought. For example, the absence of a meteorite landing on my

head is a cause of this thesis.

3. Problems specific to causal theory. Eg Dowe's analysis of causation means that causation can only happen when a quantity of energy or momentum is conserved, and an absence cannot instantiate a quantity of energy / momentum.

4. Problems specific to the causal relata. This worry is based on the idea that if the causal relata is x, absence-x either can't exist, or if it does exist, can't have causal efficacy.

5. Problems specific to causation being a relation. That is, the idea that one of the relata's being missing is a problem for thinking of causation as a relation.

6. Problems relating to disjunctivism. This is Lewis's worry: if causation is all about biff, then how can we characterise absence causation in terms of biff without our analysis becoming disjunctive.

7. Problems due to possible causation ("would-cause" counterfactuals). (Dowe 2009)

Let us approach these in reverse order. With regard to problems due to possible causation, my response is that we should accept that absence causation is derivative (or more specifically, that absence causation is only possible for the more derivative kind of causation, causal difference making).

In the case of problems relating to disjunctivism, I have argued that this depends on the details of the account. Lewis is right that the most obvious

account of causation in terms of biff will end up being disjunctive in order to account for absence causation (Lewis 2004b, pp. 284-5). There is no reason, though, to believe that a more sophisticated account couldn't do a better job; see my final chapter for a suggestion from Strevens (2013).

Another worry Lewis raises is that if causation is a relation, and one of the relata is an absence, then we have problem as one of the relata has 'gone missing' (Lewis 2004b). I think it is a mistake to phrase this in terms of a relatum being missing. Instead we should say that one (or more) of the relata is negative. The question for us then is how to understand this negative relatum so that it is non-problematic. Some problems regarding absence causation are due to our choice of causal relata. Absence events are especially problematic. My suggestion is negative facts.

Problems specific to causal theories will be discussed in subsequent chapters: 4, 5, 6, 7 and 10.

I do not have space to cover the problem of multiplying causes thoroughly in this thesis, but see chapter 11 for suggestions from Strevens (2013) and Handfield et al. (2008).

Part II

Reductive accounts of causation

Chapter 4

Lewis

4.1 Introduction

David Lewis produced two theories of causation, which I will refer to as his early view and his late view. Both are counterfactual accounts of causation, and both are arrived at by conceptual analysis. The difference is that Lewis's early view (Lewis 1973b) talks about steps of causal dependence between the cause and the effect. E is causally dependent on C iff (1) if C had not happened, E would not have happened and (2) E happens. By contrast Lewis's late view is about steps of causal influence. Instead of the question being *whether* E happens counterfactually depends on C, we can also ask if *how* and *when* E happens are counterfactually dependent on C. If so, C has influenced E. In both cases we talk about steps of influence or dependence; this ensures transitivity, since counterfactual dependence is not transitive, but in Lewis's view(s), causation is. In this chapter, I will first describe Lewis's early view, followed by his late view.

4.2   Lewis's Early View

Lewis's early view arose as a response to worries about regularity theories of causation. Lewis writes that while it is possible that there might be a version of the regularity theory which deals with all the objections, in his view it is time to try out an account along different lines: a counterfactual account.

He writes that philosophers have often shied away from using counterfactuals because counterfactuals lack clear truth conditions. Lewis's account involves providing some truth conditions to do with counterfactuals, involving similarities between possible worlds. But he also writes that we shouldn't have too high a standard about what counts as understanding counterfactuals, or how clear our truth conditions for them should be. Counterfactuals have an element of vagueness, and our account of them should reflect this. There is no getting away from this, but it is not necessarily a disadvantage.

Lewis confines his early account to particular events, to defining what it is to be a cause, not *the* cause, and only aims to cope with deterministic cases of causation. To Lewis, two possible worlds are deterministic if, as long as they share laws of nature and are the same in every particular up until a certain point, they will go on being the same for all other times as well.

To base an account on counterfactuals, Lewis needs an account of what it takes for a counterfactual to be true. Lewis does this in terms of similarity relations between possible worlds. If I say 'if I had eaten dodgy takeaway last night, I would now feel ill', then my claim is true iff one of the following is true:

1. There is no possible world in which I had dodgy takeaway last night;

or

2. There is a possible world in which I had dodgy takeaway last night, in which I now feel ill, which is closer to the actual world than any world in which I had dodgy takeaway and do not now feel ill.

(This is what it takes for a counterfactual to be true in this actual world. For a counterfactual to be true in some other world, we just substitute that world for the actual world in the definition above.)

To make sense of this, we obviously need to know what counts as a world being more similar, or closer. Lewis gives a more detailed account of this elsewhere (Lewis 1986), but broadly, how similar worlds are to each other depends on how they differ in terms of particulars and laws of nature, and a small difference in particulars counts as less of a difference than a small difference in laws of nature. The actual world is always the closest to itself, and it is possible for two different worlds to be the 'same distance' from the actual world.

The gist of Lewis's account is that we can analyse causation in terms of causal dependence, and causal dependence in terms of counterfactual dependence. The account of counterfactual dependence is the story above about similarities between possible world. Lewis takes possible worlds to be real, and similarity between worlds to be primitive. Thus his account of causation is ultimately reductive.

To Lewis, causation must be transitive. Counterfactual dependence is not transitive, so to secure transitivity Lewis needs to do more than say that A causes B iff $\sim A$ then $\sim B$. Lewis writes that when we have a case of

counterfactual dependence we have a case of causal dependence. But in order to get A causes B we don't need A to causally depend on B. We just need A to be linked to B via a chain of events, each of which is causally dependent on the next. Thus causation is the ancestral of causal dependence.

The idea of an ancestral of a relation is this: for a relation R, the ancestral of R stands to R as 'ancestor' stands to 'parent'. For any relation, if we take the ancestral of that relation, then the ancestral will be transitive: if *Rab* and *Rbc* then *Rac*. [1].

Lewis's views on causation do not depend on his ontology of possible worlds. Lewis's commitment to realism about possible worlds is unusual. It would be very possible to agree with Lewis's views about causation without accepting his views on possible worlds. Lewis's views on possible worlds do two important jobs here. They give us a story about what grounds counterfactuals and modality: possible things are made of the same stuff as actual things, but are in another world which is spatiotemporally discontinous with our own (Lewis 1986) on the plurality of worlds. And Lewis's theory of similarity relations between possible worlds provides us with truth conditions for counterfactuals. If Lewis's view of possible worlds does not appeal, we would have to find something else to do these jobs.

There is no space here to give a real survey of approaches to possible worlds. Notable alternatives, however, include ersatzism and fictionalism. Broadly speaking, Ersatzism is the view that possible worlds are maximally consistent sets of propositions (Stalnaker 1976) or states of affairs (Plantinga 1976). Fictionalism is the view that possible worlds are convenient fictions,

---

[1] This idea is due to Frege (1879).

and so statements such as 'there is a possible world where I watered the plants' is true within the fiction of possible worlds. Varieties of this view are endorsed by Rosen (1990) and Armstrong (1989).

4.2.1 Problems with Lewis's early view, and his response

The thought behind counterfactual accounts is that it is very natural to say that the truth of claims like 'eating that dodgy takeaway last night caused my illness today' depends on the truth of claims like 'if I had not eaten that dodgy takeaway last night, I would not have fallen ill today'. Lewis is right that this has a lot of intuitive appeal; indeed, if I made the second of those two claims, many people would take it that I was implying the second one.

However, there is a problem. Suppose eating that takeaway deterministically caused my illness (take whatever bodily processes are involved in (in)digestion, and just assume that they are deterministic). Then it also seems true to say 'if I had not fallen ill today, I would not have eaten dodgy takeaway last night'. These cases are known as backtracking counterfactuals.

Lewis's attitude is that it is best to just deny that this counterfactual is true. (This is also Hall's approach (Hall 2004)). It does sound as though there is something wrong with it; it is hard to imagine a context in which someone would make a claim in that form, and the claim sounds more evidential than causal. Perhaps it seems better to say that if I had not fallen ill today, I would have eaten a different, less dodgy, takeaway. Lewis's account of similarities between possible worlds supports this, since the world in which I

still have takeaway but (for example) it is freshly cooked rather than involving last week's dubious egg-fried rice is closer to the actual world than a world in which I don't have takeaway at all. (Further still, though, would be a world in which I just have stronger digestion, since that involves a more widespread difference in particulars.) Lewis argues, though, that we should go for the smallest possible change. The closest possible world to the actual one (in which I ate the takeaway and fell ill) was the world in which I eat the takeaway, and then a very small miracle occurs which prevents me from getting ill. This is a smaller change than my eating a different takeaway and, Lewis argues, definitely a smaller change than altering the laws of nature so that (for example) bacteria in improperly stored food don't make people ill. Thus, Lewis argues, it is just not true that 'if I had not fallen ill today, I would not have eaten takeaway last night'.

Another issue, which Lewis deals with in a similar way, is the problem of epiphenomena. Suppose that I get takeaway and it makes the house reek. Later, I fall ill. For the sake of argument, we can say that getting takeaway deterministically causes the house to reek. The worry arises if the following counterfactual is true 'if the house did not reek, I would not have fallen ill today'. We would not accept the causal claim that the reeking house caused my illness. It is clear that getting takeaway was the common cause. Notably, we could not even fix this by saying that the cause must come before the effect. That approach might have fixed the problem above, of reverse counterfactuals, but it doesn't solve the problem here, since the smell comes before the illness. Lewis does not want to rule out cases of the effect coming before the cause a priori, since this seems like it should be an empirical

matter. But when it comes to the problem of epiphenomena, it wouldn't help in any case. The only solution seems to be to deny that the counterfactual is true.

If I had not fallen ill today, it is not the case that I would not have eaten the takeaway. Thus, the house would not reek. Again, the closest possible world is one in which I eat the takeaway, it makes the house reek, everything else proceeds as normal, but some small miracle prevents my getting ill.

4.2.2 Problems with Lewis's early view which remain

A major problem with Lewis's early view is that of pre-emption. Lewis discusses this in the final section of his 1973 paper 'Causation'.

Here is one type of pre-emption, early pre-emption: we have two possible causal chains leading up to the effect e. a causes b, which causes e. c causes d which also causes e. c also prevents b, thus breaking the causal chain from a through b to e. In these sorts of scenarios, c is clearly a cause of e. However, for some kinds of counterfactual account, it won't be, because if c had not happened e would still have come about. (Paul and Hall 2013, p. 75).

Lewis's counterargument runs as follows:

> As far as causal dependence goes, there is no difference: e depends neither on c_1 nor on c_2. If either one had not occurred, the other would have sufficed to cause e. So the difference to must be that, thanks to c_1, there is no causal chain from c_2 to e; whereas there is a causal chain of two or more steps from c_1 to e. Assume for simplicity that two steps are enough.Then e depends causally on

some intermediate event d, and d in turn depends on c_1. Causal
dependence is here intransitive: c_1 causes e via d even though e
would still have occurred without c_1.

(Lewis 1973a)

The idea goes like this. We can distinguish which of c or a is the real
cause by tracing the chain of causal dependence. The effect does not just
need to counterfactually depend on the cause, but each step in the chain
between cause and effect needs to be a case of causal dependence.

This kind of example is known as early pre-emption. The trouble is that
Lewis's response does not adequately deal with another kind of pre-emption:
late pre-emption. This happens when for example, the takeaway is so dodgy
that there are two kinds of food poisoning bacteria in it (call them type 1
and type 2). As it happens, the type 1 bacteria are more plentiful, so they
cause my nausea first. But if they hadn't caused my nausea, then the type
2 bacteria would have done this too, if a little later. There is no obvious
different causal chain here, so Lewis's response to the early pre-emption case
in his 1973 paper does not work here.

The other thing which is notably absent from Lewis's early view of causa-
tion is any treatment of causation by absences. What if my nausea is caused
by having had nothing to eat all day (perhaps all this talk of food poison-
ing has made me paranoid)? Then it seems as though our causal relation is
missing a relatum. As we will see in the next section, this is a problem which
Lewis's late view aims to fix.

4.3 Lewis's Late View

In his papers 'Causation as Influence' (Lewis 2004a, pp. 75-105) and 'Void and Object' (Lewis 2004b), Lewis revised his earlier view, to cope with two worries: late pre-emption and causation by absences.

There are two kinds of late pre-emption, both worrying.

Here is an example of trumping pre-emption (the term is due to Schaffer (2000)). Suppose I am climbing a mountain and pause to consider my route: should I go left or right? I have a map with me, and I also have a GPS device. If the GPS and the map disagree, I will decide my route based on the GPS, because I know its electronic map has been updated more recently than the paper map I also have with me. However, in this case, they don't disagree; both indicate that we should go left. So we do. If I did not have the GPS with me, the map would still have made me turn left. But as it is, as far as Lewis is concerned, we want to say that my going left is caused by the GPS not the map. This looks like a worry for Lewis, since there is a case where we want to say there is causation without counterfactual dependence.

The other kind of pre-emption which represents a problem for Lewis's early view is late pre-emption. Suppose two people from opposite ends of the building both smell smoke at the same time. Both realise there is a fire, and both hit the fire alarm buttons nearest to them, respectively. They hit the buttons at nearly the same time but Annie gets there a split second before Becky, meaning that it is Annie hitting the fire alarm button near her that sets the alarm off, not Becky hitting her nearest fire alarm button. However, even if Annie hadn't hit the button, Becky would still have pressed

the button near her, so the fire alarm would still have gone off.

4.4   How the idea of influence allows Lewis to respond to critics

Swapping out the idea of causal dependence for the idea of causal influence allows Lewis to respond to these worries about the counterfactual approach: specifically, late pre-emption and trumping.

Lewis's new account of causation as influence deals with the case of late pre-emption like this. In his early view, Lewis took it that counterfactual dependence must be 'whether-whether' dependence. Roughly, if the cause had not happened, the effect would not have happened. But in his late view, he argues that this could just as well be when-on-whether dependence, or even in-what-manner on whether dependence. To take our example of Annie and Becky and the fire alarm, *whether* the fire alarm went off does not counterfactually depend on Annie's hitting the alarm button, since Becky hits one too just after. But *when* the alarm goes off does counterfactually depend on Annie's pressing the button, since Becky hit the alarm a split second later and it would have sounded slightly later than it in fact did, if Annie had not pressed the button.

In the case the case of the *manner* of the effect counterfactually depending on the cause, we can consider the case where Annie and Becky both smell smoke and Annie hits the fire alarm. But Becky's instinctive response is to pick up the phone and call the fire brigade. Say the building is a hospital, so

the alarm system automatically calls the fire brigade if the alarm is set off. So whether Annie had hit the fire alarm or not, the fire brigade would have been alerted. But if Annie had not hit the fire alarm, they would have been alerted in a different way (by Becky phoning them). Lewis describes this as not causal dependence but causal influence, and it plays the same role in his late view as causal dependence did in his early view.

In the case of trumping, the idea of influence is also helpful. In the case of the map and the GPS, if I am turning left because of the GPS, not the map, then changing something about the GPS would make much more of a difference to my action than changing something about the map. If the GPS turns out to be malfunctioning, and this leads to it giving me poor directions, then I might turn right instead of left. On the other hand, if the map dates from the 1920s this will makes no difference to my actions.

4.4.1 Causation by absences

Another aspect of Lewis's late view is that he tackles the question of causation by absences. The issue of whether absences can be causes (for example, if the lack of warmth causes my shivering) and whether they can be effects (for example, if the frost caused the lack of strawberries) divides philosophers of causation. It tends to be a good litmus test for what they take an analysis of causation to be. If they take analysing causation to be a process of finding out what is going on in the world, they tend to rule out causation by absences because, after all, absences are not things in the world. On the other hand if they take themselves to be analysing what it is we are referring to when

we talk about causation, they are likely to allow causation by absences. This is because we do after all talk about absences as though they can be causes and effects (I discuss this in more detail in chapter three above).

Lewis's approach to causation is conceptual analysis so it is perhaps not surprising that he takes it that causation by absences is possible. However, he warns that anyone who thinks that there is causation by absences will still have to cope with the challenge that this means causation is going on all the time: for example the absence of scorching heat is causing my skin to not blister. Meanwhile anyone who thinks there that there can be no causation by absences has to answer the question of what happens when the absence of warmth in the dry ice causes my frostbite when I touch it. In other words, Lewis's point is that our intuitions do not necessarily support causation by absences: rather, we support some causation by absences and not others.

4.5   Lewis and biff

In much the same way as Dowe accepts conceptual accounts of causation as a separate enterprise from his own, Lewis argues that accounts of causation such as Menzies' are good accounts, but just not accounts of causation. Instead, he says, Menzies has given an account of Biff. Not all causes are related to their events by means of biff, but causal relationships can be described in terms of biff, or absences of it. Biff is an intrinsic relation, but there can still be cases where causation is an extrinsic relation. (Roughly, relation is intrinsic if it holds in virtue of something about its relata; it is extrinsic if it depends on things apart from the relata. I discuss this in more

detail in chapter 6 below.)

The idea is that we define causation counterfactually, and make it broad enough to include the various different kinds of causation, including double prevention, causation by absences and so on, whereas biff is narrowly defined and can be approached empirically. Lewis writes that the account of causation due to Fair (1979) where causation is characterised as a transfer of energy or momentum should work as an account of biff; it seems as though this is how he would see Dowe's view as well.

So Lewis's disagreement with those who advocate a physical account of causation instead of a counterfactual one is not so much a matter of his thinking that their account is wrong, but that it is not an account of *causation.* There is a parallel then between Dowe, who thinks we can have a conceptual analysis of causation alongside a different empirical analysis, and Lewis, who argues that we can have an account of biff as well as causation; both think that we may need two accounts to make sense of causation. I will discuss this further in chapters 6 and 11.

4.6 Critique

Lewis's account of causation is bound up with his possible worlds realism. According to Lewis, counterfactuals are to be understood as claims about possible worlds, and the truth conditions for counterfactuals are conditions to do with similarities between possible worlds. If we sign up to Lewis's possible worlds realism then we are more likely to find Lewis's account satisfying. Ultimately, Lewis's account defines everything in terms of particular

matters of fact and similarities between them, even though some of these are in other possible worlds. This means that everything is ultimately being explained in terms of something physically and metaphysically unmysterious – as long as we are happy with the existence of possible worlds in the sense Lewis is thinking of them. Causation is then being explained in terms of something which is in principle empirically accessible. However, many consider possible worlds realism to be implausible. Other ways for understanding possible worlds, such as ersatzism or fictionalism, make possible worlds mind-dependent. Reducing causation to counterfactuals, understood as claims about the similarity relations between possible worlds, and reducing those to something mind-dependent, means that we are ultimately saying causation is mind-dependent. This is unpalatable.

Both Lewis and Dowe acknowledge that two analyses are needed. For Dowe, this is the conceptual analysis of causation and the empirical analysis of causation. For Lewis, there is causation and then there is biff. This raises the question: why doesn't Lewis call biff 'causation'? It seems fairly clear that biff relations, both those that occur in this world and those which occur in other possible worlds, are what makes causation in Lewis's sense possible at all. Events are only able to make a difference to other events because there are biff-relations between them. Lewis leaves open the possibility that other worlds might have causation, but with something else playing the role of biff-relations, but it is hard to see what that might be.

Absences as causes and effects present a puzzle. Lewis argues that the solution here is to deny that causation is a relation, and so we don't have to worry about a 'missing' relatum. I will discuss this in more detail in

chapter 6, where I discuss whether the Canberra Plan can be used to analyse causation. Here, however, I want to discuss the problems which can arise if we accept both the thesis that causation is transitive and that absences can be causes and effects. A classic worry about absences and transitivity goes like this (Lewis 2004a, p. 96): there is a forest fire in June. One of the causes of the fire is the presence of flammable material, which is only present because of the lack of a forest fire in May. The lack of a forest fire in May is caused by the rain. So the rain in May causes the forest fire in June: a very implausible idea. We will only count this as causation if we accept that causation is transitive, and that the lack of a forest fire in May can be a cause. On the other hand, both theses about causation do seem plausible. This is obviously a problem for Lewis's early view; but his later view seems to avoid it. If we understand causation as influence, then when and in what way something happens can be causes as well as *that* it does happen. In this case, it sounds much more plausible to say that the rain's being in May rather than June caused the forest fire in June. In chapter 3 above I discuss the idea that causation might be constrastive: that is, that it might be a four place relation between cause, effect, contrast to the cause and contrast to the effect. Another possibility which I found more convincing was the notion that the causal relata are facts eg 'that there was a forest fire in June', rather than events eg 'the forest fire' (which occurred in June). In some ways these approach are the natural extension of Lewis's causal influence approach.

4.7   Conclusion

Lewis's theory is built around the basic idea that for c to cause e is for it to be the case that if c had not happened then e would not have happened, developed to fulfil our intuitions (for example, taking the ancestral to make it transitive) and address various worries. According to Lewis's late account, when and how something happens are as crucial as whether it happens. Causal claims for Lewis are ultimately to be understood in terms of counterfactual claims, which in turn can be understood in terms of similarity relations between possible worlds.

Problems for Lewis's early account include trumping pre-emption and late pre-emption. His late account uses the notion of causal influence to deal with these worries. Lewis's account of causation is reductive. How far this reduction goes depends on what view we take of possible worlds. If we signed up to possible worlds realism then Lewis's account is fully reductive, but other views of possible worlds such as ersatzism or fictionalism make them mind dependent. This limits how reductive Lewis's account can be.

Chapter 5

Dowe

5.1 Introduction

In this chapter I will describe the account of causation offered by Dowe in his book 'Physical Causation' (Dowe 2000). I will argue that his account is successful but that there are things which we need from our account of causation which Dowe's account does not cover, and which we will need a second account to handle. In other words I am arguing that causation as understood by Dowe is one of the two kinds of causation.

Dowe's account of causation is based around ideas of causal processes and causal interactions. Dowe first defines an object (using a very permissive definition) and then a process. He then specifies criteria by which we can tell which processes are causal. Causal interactions are then defined in terms of causal processes. Roughly speaking, a process is causal when a quantity of energy or momentum has been conserved, and a causal interaction takes place when a quantity of energy or momentum is exchanged.

The benefit of this approach is that energy, momentum, objects and processes are all real, physical, concrete things which we can understand without an appeal to causal concepts. Dowe's analysis, then, represents a reduction of causation to something non-causal. Even more importantly, those non-causal elements of the analysis are empirically accessible. This explains how it is possible that we know anything about causation (I will describe in more detail how I think it explains that below in chapter 9).

After describing Dowe's account I will go on to consider some worries. Firstly, I will look at the concern that Dowe's account of causation is too much unlike our every day concept, and Dowe is therefore changing the subject. Secondly, I will discuss the worry that in places Dowe does rely on intuitions and linguistic concerns for his arguments. Thirdly, I will look at the worry that Dowe's account does not deal adequately with causation by and of absences. One point which needs to be made here is that especially the first and the third criticism are in some ways forestalled by Dowe, in his labelling of his account as an empirical analysis. His response to these sorts of criticisms is that they are criticising his account for not doing something that it is never intended to do in the first place. With this is mind, my comments should be taken as criticisms of the general empirical analysis approach, at least unsupplemented by some other account which makes sense of absences causation. For example, perhaps an empirical analysis of causation does not have to present a concept which works in the case of causal explanation. But then, that is a problem for empirical analysis when it comes to causation, since making sense of causal explanation is an important desideratum.

Finally, I will conclude that Dowe's account of causation is successful as

an account of one kind of causation, but that we need a second account to explain the other kind.

5.2 Dowe's theory of causation

In this section I will give a brief outline of Dowe's theory of causation. Dowe's theory belongs to a family of causal theories know as process theories of causation. Other theories in this category include those due to Salmon (1984) and Fair (1979).

Dowe's account of causation is based on causal processes and causal interactions. What causes and effects are, what distinguishes them from each other and what the relation between them is, can then be explained in terms of causal processes and interactions.

Often when philosophers are trying to analyse causation they are aiming to analyse a causal relation between distinct relata (usually events, though see my introductory chapter on causal relata). This is roughly what Dowe is talking about when he refers to causal interactions. Dowe also talks about causal processes however. This captures his belief that a process can be causal in the sense that persistence and inertia are causal. I will discuss this in more detail below.

To Dowe, causation is the exchange of a conserved quantity such as charge or mass-energy. This happens when the world line of one object possessing a conserved quantity intersects another. Two things with conserved quantity meet each other, some conserved quantity is exchanged; causation has happened.

Dowe writes:

> CQ1. A *causal process* is a world line of an object that possesses a conserved quantity. CQ2. A *causal interaction* is an intersection of world lines that involves exchange of a conserved quantity. (Dowe 2000, p. 90)

Obviously there are a lot of terms here which need explaining. A world line is 'the collection of points on a space time diagram that represents the history of an object' (Dowe 2000, p. 90). In other words, we can imagine an object drawing a line behind itself as it travels through time and space. When the object which is travelling possesses a conserved quantity, we can see that line as a causal process.

A conserved quantity is 'any quantity that is governed by a conservation law'. Examples are mass-energy, linear momentum and charge. Dowe takes our current scientific theories to be the best guide to what these quantities are.

Dowe takes a very permissive view of what counts as an object: 'anything found in the ontology of science (such as particles, waves or fields), or common sense (such as chairs, buildings and people)' (Dowe 2000, p. 90). He even includes spots and shadows. Just being an object doesn't make something causal. The concept of an object isn't really doing a lot of work in Dowe's theory, in terms of understanding what causation is or for distinguishing the causal from the non-causal[1].

---

[1]The only kind of thing which Dowe will not allow to be an object is what he calls 'time-wise gerrymanders'. A gerrymander is something like an object but without any real unity. If it is a random collection in space that would be a spatial gerrymander, while if

Finally, we need to understand what Dowe means by 'possess': he means 'instantiate'. The object possesses charge, for example, if we can truly say that the object is charged. This is as opposed to the idea of the conserved quantity being a separate thing which somehow gets transmitted, or where the quantity at the end of the process is the 'same' as what you started with, in the way as this is the same laptop which I used this morning.

Dowe's account is singularist. That is, the particular case of an individual object traveling through space-time is more fundamental than general causation between types of thing. Dowe argues that this is the case even though he appeals to general theories about which quantities are conserved. What matters is that the particular quantity possessed by the particular object whose world line we are talking about, is conserved. The general scientific theory is just how we know that the quantity is conserved, because to the best of our knowledge the quantity is of the kind which is conserved. Whether a certain process is causal does not depend on what is going on elsewhere in the universe, either in time or space: the causal relation, as characterised by Dowe, is intrinsic. To many writers (for example, Menzies (1996), Hall (2004)) this is a desirable quality for an account of causation to have.

Dowe's account allows for the universe to be indeterministic. Because a causal process is a particular process, there is no idea of necessity built into the theory (Dowe 2000, p. 113). It is all right that an atom of an unstable

---

there is not unity through time either, then it is a time-wise gerrymander. An example, not the spot of light on a wall (Dowe is happy for that to be an object) but the patches of wall which are illuminated, one after the other, each at different times but thought of together. This does not persist through time; different bits of it persist at different times. To Dowe, these are not objects, and if they have world-lines, those lines are just spatiotemporal junk.

element might or might not decay, and that whether and when it happens is irreducibly indeterministic. If the atom decays, then the charged particles it gives out are involved in causal processes. But if the atom does not decay, that causal process does not happen. Dowe's account avoids counterfactual elements, and so avoids puzzling aspects of indeterminism, and is also supposed to avoid issues of overdetermination.

Dowe's account allows for there to be a lot going on in the universe which is not causal. In other words, Dowe's account involves rejecting the principle of causality, that every event has a cause. Few philosophers in modern times accept this 'law' of causality in any case.

5.3   Dowe's empirical analysis

5.3.1   Empirical analysis as analysis of the concept inherent in scientific theories

Dowe describes his account of causation as empirical analysis. He distinguishes two approaches to analysing causation: empirical analysis and conceptual analysis. In his view, conceptual analysis consists in analysing our concept of causation, whereas empirical analysis consists in analysing what causation is in reality.

Dowe argues that it is important not to criticise accounts of causation for something the writer never intended it to be in the first place. He cites examples where empirical accounts have been criticised for not matching our ordinary ideas of causation. The point is, he argues, we wouldn't expect them

to; an empirical account of causation does not have to track every aspect of how we use the word 'causation'. Dowe's empirical analysis is also contigent and intentionally so, so he argues that possible worlds based criticism will be misplaced. I discuss this further below.

Dowe writes that we need, as he puts it, 'horses for courses'. We need more than one definition of causation to deal with the variety of demands we make on a definition. An empirical theory is suppose to tell us what causation really is in the world, while a conceptual analysis gives us an account of our common sense use of the word. Neither type of definition should be expected to do the job of the other. Dowe is aiming for an empirical analysis. (This kind of analysis is also described as ontological, as opposed to conceptual, since it is to do with what actually exists. (Mellor 1995))

He also adds that empirical analysis, like conceptual analysis, has the potential to be revisionary, in the sense of leading us to revise our common sense idea of causation. We might find that our empirical analysis and our everyday view match up, or we might feel that it doesn't matter that they don't. On the other hand, we might feel that we should change our everyday concept to match our empirical analysis.

The obvious objection here is that Dowe may end up with an analysis of something which is not causation. If the analysis doesn't have to describe something which matches up with our everyday concept of causation, then Dowe could be accused of changing the subject.

Dowe's answer is simply that we are deriving our empirical analysis from scientific theories, and scientists are after all competent users of the word 'causation'. The same word, being used in the same sense, is enough. Dowe

considers whether empirical analysis is the same as a conceptual analysis of how scientists use the word in question. He writes that, certainly for his purposes, it is not. Dowe's analysis could however be classified as a conceptual analysis of how the concept of causation is used in science, although not necessarily how it is used in the 'linguistic practices of science' where those differ from standard English usage.

5.3.2   Contingent analysis

Dowe describes several different kinds of empirical analysis. One kind of empirical analysis could give us a contingent truth about causation, while another kind would give us an a posteriori necessity (in the same way that Kripke (1980) regards 'water is H_2O' as an a posteriori necessity). Dowe is aiming for the first of these options.

Dowe's account of causation is contingent. Dowe writes that it should hold true in any physically possible world. But any possible world with different laws of nature will work differently, and Dowe's account of causation will not necessarily hold true. This means that what causation is not only something to be establish a posteriori, but is also something which might have been different if the laws of nature had been different.

5.4   Dowe's analysis of causation: worries

In this section I will look at three worries about Dowe's approach to analysing causation. One concern is that he is changing the subject. That is, because his analysis has little to do with our concept of causation in everyday life,

and because he is not concerned about objections based on our intuitions about cases of causation, Dowe's analysis risks not being about causation at all.

A second, related, worry is that in fact Dowe does use appeals to intuition in several places in his argument. My point here is that whilst we can have an empirical analysis of causation, we can't have an analysis which is completely non-conceptual.

My third worry is absences. Dowe's account of causation does not allow for causation by or of absences. He writes that if we rephrase our statements involving causation by or of absences we can get some more accurate description of the situation which does not not involve absences as causes and effects. I will argue that this response is not successful and Dowe's account of causation fails to do justice to causation by and of absences.

5.4.1 Changing the subject

A basic worry with any revisionary analysis of causation is the risk that we simply drift so far from our everyday conception that we are no longer really talking about causation. There are many ways in which we talk about cause on an everyday basis which do not seem to be covered by Dowe's theory. For example, we might say 'the sunshine made me happy'. The sunshine and my happiness do not seem to be linked by a causal process in the way Dowe defines it, even though one causes the other. There doesn't seem to be a transfer of energy or charge involved, or an object persisting in having a quantity like that. If we think that everything is fundamentally explicable

in physical terms then presumably the causal link between the sunshine and my happiness is physical. With a complete physical theory, we might even be able to trace the causal chain. But we would not normally think of this causal connection as a complicated one. Under a difference-making (counterfactual) theory, it is quite simple.

Dowe argues that his account of causation is not meant to reflect ordinary usage. So this kind of criticism has no weight for him. But he does acknowledge the need for there to be some kind of link between our every day concept and his analysis.

Dowe argues that this link is supplied by the historical connection. He writes that there is clearly a connection between how 'cause' is used in the context of science and how it is used in every day contexts. In scientific contexts, 'cause' is not a technical term with a specialised definition, nor is the word being used ironically – it is being used in just the same sense as usual. For Dowe, this is enough of a connection (Dowe 2000, p. 9).

Although there is a connection between Dowe's analysis of causation and our everyday concept of causation, there are still many aspects of our concept of causation which Dowe's account does not do justice to. Woodward (2003) criticises Salmon's account (which for all relevant purposes here is similar to Dowe's) on the grounds that our understanding of causation is what gives us explanations, and the kinds of explanations we get from a process account are the wrong ones. For example, when we want to explain why the air from a compressed air canister is cold, we do not explain it on the basis of the energy and movements of each individual particles. Rather, we appeal to laws about the behaviour of gases in general, and facts about the situation

at a macroscopic level.

This is not a reason to reject Dowe's account or a problem with his account in itself. Rather it is a limitation of his approach. It provides an account of how causation can be grounded at the fundamental level, but does not explain how we use our concept of causation. Dowe says as much in his introduction when he suggests that his account is an empirical account, but a conceptual account might also be needed.

5.4.2 Dowe's use of intuitions

A related but slightly different worry about Dowe's account is that, in places, it does seem as though he is appealing to intuitions. If intuitions are only useful in clarifying our everyday concept, and not for finding out what causation is in reality, then Dowe's use of intuitions will be incongruous with his stated aims.

For example, Dowe defends his commitment to causation as indeterministic by appealing to 'everyday talk' and 'scientists' talk'. Dowe's argument goes like this:

1. The connection between C and E is a causal connection.

2. The probability P(E|C) = x is irreducible.

3. Therefore, not all causes are sufficient conditions for their effects.

(Dowe 2000, p. 23)

The reason for believing (2) is that science tells us that there are some events which are connected in a way that is irreducibly indeterministic.

Dowe's two arguments for (1) however ar based on intuitions as well as science. The first argument is very simple: the decay of radio-active material is indeterministic. Yet, fairly obviously, if I bring some dangerously radio-active material into a room, and gives you radiation sickness, then I am responsible for your sickness. I cannot be held responsible for something unless I caused it or at least failed to prevent it, so the process of radio-active decay is causal, even though it is indeterministic.

This is an appeal to intuitions if there ever was one, and Dowe recognises this:

> It may be objected here that I am making an illicit appeal to
> everyday intuitions about meaning – illicit since I am using them
> to criticise Hume's empirical theory and since I myself am seeking
> to develop an empirical theory.
> (Dowe 2000, p. 23)

Dowe's response is to say that it is fine to use cases and appeals to intuitions to shake preconceptions. He argues that the view that causation can only be deterministic has no basis in science, and so a case which makes us question determinism about causation further is a helpful tool.

Dowe's second argument is just that scientists call the decay of radio-active materials 'production', and production means the same thing as causation. In other words, scientists do not see any problem with something being causal, even when there is no necessity and no deterministic causation. Dowe admits that this isn't a conclusive argument, but writes that if those who think causation must always be deterministic want to disagree

that would be 'an adhoc response' (Dowe 2000, p. 24).

So the role of both our ordinary talk and scientists' talk here is to shift the burden of proof. Instead of Dowe needing to say how something could possibly be causal without it being deterministic, it is up to those who think causation is always deterministic to say why indeterministic sequences of events shouldn't be causal, given how we normally talk about them as causal.

Dowe writes that his account of causation is meant to describe the concept of causation emerging from science, not necessarily reflect the linguistic practices of scientists. However, in one place it does look like he is appealing to the linguistic practices of scientists to make a key point in his argument.

Dowe argues that inertia should be seen as causal, just as much as interactions. That is, he argues that a rock in space which keeps moving under its own momentum is just as much a causal process as a one rock hitting another and making it change direction. One way he argues for this is to say that the scientific definition of inertia is 'the property of a body in virtue of which...' (Dowe 2000, p. 54) and 'in virtue of which' sounds like a synonym for causes. That is, according to the scientific definition of inertia it is a causal thing, so our account of causation should include inertia as a kind of causation.

This seems a lot like drawing a conclusion about causation based on the linguistic practices of scientists. It is not clear how the idea that inertia is causal emerges from scientific theories, except in that the definition of the word suggests that it is causal.

So Dowe's account is meant to be an empirical analysis of causation as it is in reality, not a conceptual analysis based on our intuitions or linguistic

practices. However, in places, Dowe does base arguments on intuitions and his observations of linguistic practices.

However, these sorts of criticisms, for example the worry that Dowe's account does not work for causation by and of absences (I discuss this in the next section), do not form a reason to reject Dowe's account. They merely show that his account is not the whole story.

5.4.3   Absences

In chapter three I discuss why, in general terms, causation by and of absences is often a problem for accounts of causation. Most accounts find some way of accommodating it. Dowe, however, takes a different approach. He argues that there is no causation by and of absences, and there only appears to be due to sloppy phrasing on our part. For example, it seems like my failing to stir the soup causing it to burn is a case of an absence causing an effect. But according to Dowe this just appears to be this way due to how the situation is described. A better way to describe it would be to say that particles of soup near the bottom of the pan remained motionless relative to the heat source, and burned for that reason. Dowe argues that this is not so unintuitive; once we have appropriately rephrased examples we are inclined to say that this is a better description of what happened than describing it in terms of absences.

In 'Absences, Possible Causation and the Problem of Non-Locality' Dowe (2009) argues against absence causation for two reasons, which are both motivated separately from the details of his account. His worries are that

absence causation is a problem for Actual-Possible independence, and that absence causality violates locality.

I have discussed the issue of A-P independence in chapter 3 above. My suggestion is that we accept that absence causation is derivative, or rather that it is something that only occurs in the context of causal difference making, which is the more derivative form of causation.

The issue of locality will be discussed in more detail in chapter 10, where I discuss Hall's suggestion that there are two concepts of causation. According to Hall, three theses about causation: locality, transitivity and intrinsicness, clash with two other theses: omissions and dependence. The solution is that there are two concepts of causation. I will also discuss Dowe's worries about Hall's approach in this chapter.

In his paper, Dowe (2009) does contemplate what he calls a 'two varieties' approach to causation, which is similar to the two-tiered approach I suggest in chapter 11. His view on this is that the two varieties have so little in common that calling them both 'causation' is a stretch. To Dowe, absence causation is merely quasi-causation. I will discuss Dowe's view in chapter 11, where I will argue that it makes sense to call both varieties causation.

My view is that the thesis that causation by and of absences is possible is such a crucial part of our concept of causation that denying it is too high a cost. There are too many cases where denying absence causation leads to an absurdity:

For example, in 'Void and Object' Lewis (2004b) describes the void as a region empty of any particles, fields, forces or energy: a region which is utterly empty. He writes that if a person were cast into the void they would die.

Since there is nothing at all in the void, the obvious candidate for the cause is an absence. It is hard to rephrase this in terms of non-absences. One of the fatal things would be depressurisation. The person's blood would boil. There does not seem to be any cause for this besides the absence of atmospheric pressure, or if you like, the absences of particles in the atmosphere colliding with particles in the person's body. This is a fairly outlandish example. So let us take a case which has certainly happened in reality: beheading causing death. The axe slicing off the head causes death, but indirectly via some events which can only be described in terms of absences, such as the blood flow to the brain stopping. If these cases are not causation, that means that some physically unmysterious apparent obvious cases of causation are non-causal, and that sounds like a bad result.

Moreover, this case seems like one which is in the spirit of Dowe's account. The kind of causation we are talking about, physical causation between singular, concrete events, is the sort of causation Dowe's account is meant to handle. The only issue is that in this case it is not the transfer of energy which is causal, but the fact that energy is not being transferred any more. Dowe's approach to absence causation is a weakness of his account.

5.5   Conclusion: Dowe's account of causation as one kind of causation

Below in chapter 9 I will argue that there are two accounts of causation. Here, I want to argue that Dowe's account is a good candidate for being one

of those kinds.

Dowe's account of causation is an excellent example of a process account of causation. I will remain neutral on the question of whether it is better than Salmon's account or Fair's account, but it is certainly a good example of its kind. It offers a reduction of causation to physical processes and interactions.

I will argue below in chapter 11 that this kind of reduction is of epistemological value, because it explains how it is possible that we know anything about causation. To writers such as Hume causation will always be mysterious: something which we do not have epistemic access to. Dowe's account explains how epistemic access is possible.

So Dowe's account meets some of our needs when it comes to an analysis of causation. But there are also things it cannot do for us. One is provide an account of causation which makes sense in terms of causal explanation. Another is account for causation by and of absences. Below in chapter eleven I will argue that there are two kinds of causation. One is best described by a process account such as Dowe's. The other is a counterfactual, difference-making account. This second kind of account will be able to do the jobs which Dowe's account cannot do. Dowe's account does not work in the context of causal explanation; we rarely if ever give causal explanations in terms of the fundamental processes and interactions of particles. A counterfactual difference-making account is more suitable here. Similarly, Dowe's account cannot handle causation by and of absences, but this does not present a problem for a counterfactual account. In this way, Dowe's account can be one of two accounts of causation, providing some but not all of what we need from an account of causation.

Chapter 6

The Canberra Plan

6.1 Introduction

In chapter four above I discussed Lewis's conceptual analysis of causation and argued that some problems with it remain. I also suggested that there are two main approaches to causation: the empirical, biff-like approach and the counterfactual approach. When talking about causation we want to anchor our discourse to our everyday concept of causation: that is, ensure that we are actually talking about causation and not something else. On the other hand, we want to understand causation in terms of something in the external world, something physical and empirically accessible. It seems that even writers who focus on just one of these when producing their account of causation still feel the tension between these two sets of criteria. One way to try to combine these into one account is the Canberra Plan.

In this chapter I will first explain what the Canberra Plan is. I will then give a brief outline of how Menzies uses it to analyse causation, to give us an

idea of how the Canberra Plan might work in practice. Next I will look at Lewis's worry that any Canberra Plan style analysis will yield a disjunctive account of causation, as the role of causation might be filled by some other relation in another possible world. This is known as the miscellany objection. Lewis argues that we need to know what makes our concept of causation non-disjunctive.

Next, I will look at Liebesman's reply to Lewis. Liebesman (2011, p. 204) maintains that the Canberra Plan is a perfectly good approach to causation. He argues that the miscellany objection is not a problem, on the basis that our folk theory "unifies" our concept of causation. I will argue that a folk theory is not enough to unify our concept of causation in a meaningful sense.

My view is that the way to deal with the miscellany objection is to deny that the role of causation could be filled by some other relation at another possible world. Instead, I will argue that the identity of the occupier of the role is an a posteriori necessary truth.

I will conclude that the miscellany objection is not an insurmountable problem for the Canberra Plan.

6.2   The Canberra Plan

The Canberra Plan is an approach to analysis taken by David Lewis and Frank Jackson. The idea was inspired by F. P. Ramsey's paper 'Theories' (Ramsey 1990, pp. 112-136), which gives us the phrase 'Ramsey sentence'. The idea is that we can write any theory in the form of a long sentence containing all the claims of the theory joined together. This will contain

two kinds of terms: O-terms, terms in the old language which have not been introduced by the theory, and T-terms, which are new terms introduced by this particular theory. We can then remove the T-terms and replace them with a bound variable. (In Ramsey's version there is only one T-term, so we replace every instance of it with a bound variable. Lewis (1970) pointed out that it would be possible for there to be many different T-terms; in those cases, we replace instances of one T-term with one bound variable and the other with another bound variable.)

The whole sentence is then prefaced with as many existential quantifiers as there are bound variables. So, for example, if our original theory is 'genes store and transmit biological information', a rough way of paraphrasing this into a Ramsey sentence would be to say 'there is a thing, and that thing stores and transmits biological information'. The question of whether there are genes is then the question of whether the Ramsey sentence is true. If the Ramsey sentence is false then we must be eliminativists about genes. If nothing fills the role of the T-term, it looks like it doesn't exist.

There are, then, three stages to the Canberra Plan.

1. We articulate a folk theory of the concept, setting out how people generally use the term 'cause'.

2. We then take the folk theory and remove from it all the occurrences of the key word whose meaning we are trying to analyse and substitute a bound variable, to form a Ramsey sentence. We then preface the statement of the folk theory with an existential quantifier. So we have a sentence that reads something like 'There is an x, and x allows us to

bring about its effect, x is a particular event, x necessitates its effects.'

3. We then investigate empirically whether there is anything in the mind-independent world which fits this portrait of x. This empirically discovered thing, whatever it is, is x.

6.2.1   Menzies and the Canberra Plan

To see how using the Canberra Plan to analyse causation might work out in practice, let us look briefly at the account of causation due to Menzies (1996).

This approach has been applied to causation by Menzies (1996, p. 96). He suggests we use a fairly simple version of the folk theory, with three main platitudes, to form the first stage of the Canberra Plan.

Firstly, causation is a relation between distinct events. Menzies uses this to distinguish it from other kinds of dependence. He gives examples of several other kinds of dependence which are non-causal, such as the example of immunity from infection depending on a person's having antibodies in their blood stream. This is not causation because the being immune and the having anti-bodies are not distinct. For the dependence between events to be causal dependence, those events must be distinct.

The next important platitude is that the causal relation should be intrinsic to the cause and effect. In other words, whether causation takes place between a specific cause and a specific effect depends on the nature of those events and what happens between those two events. It does not depend on anything outside of those events, such as on any regularities, or by alterna-

tive processes. Menzies notes that this contradicts Hume, but that that is the nature of the commonsense conception of causation. Menzies argues that we can see this principle at work when we look at pre-emption cases. We want to say that Suzie's rock broke the window, even though the window's shattering does not counterfactually depend on Suzie's since Billy also threw a rock a split second later. The causing has intrinsically to do with Suzie's throw and Billy's is irrelevant. So, Menzies says, our concept of causation is an intrinsic relation.

Menzies offers a definition of intrinsicness in terms of the notion of a natural property. He writes 'A relation is intrinsic if it supervenes on the natural properties of its relata' (Menzies 1996, p. 99) or on the natural properties of its relata along with the natural relations holding between them. The concept of naturalness which he uses is Lewis's: natural properties are those which carve nature at the joints. If physicalism is true, then fundamental physics gives us an inventory of natural properties.

Menzies characterises intrinsic relations in terms of natural properties and relations. He defines stipulates that two objects are 'duplicates' if they have the same natural properties. He writes: 'a relation is intrinsic to its relata if and only if, whenever a and a' are duplicates and b and b' are duplicates, then both or neither of the pairs $< a, b >$ and $< a', b' >$ stand in the relation' (Menzies 1996, p. 99). This would be a way of being an intrinsic relation which supervenes on the natural properties of its relata. He argues that causation is an intrinsic relation which supervenes on both the natural properties of its relata and the natural relations between them. To do this he defines the notion of duplicate pairs thus: '$< a, b >$ and $< a', b' >$ are

duplicate pairs if and only if a and a' have the same natural properties, and so do b and b', and also the natural relations between a and b are exactly those between a' and b'.' A relation is then intrinsic to its pairs iff whenever $< a, b >$ and $< a', b' >$ are duplicates pairs, they must either both stand in the relation, or neither of them stand in that relation. Causation is, according to Menzies, intrinsic to its pairs. Whether a causal relation obtains has to do with the relata, the natural properties of the relata, and the natural relations holding between those relata, and not to do with other things going on in the world.

The third platitude is that the cause increases the chance of the effect, understood in Lewis's counterfactual sense of chance raising. Chance raising and causation do come apart in cases of overdetermination and pre-emption, and also in this case due to Hesslow (1976): birth control pills raise chances of thrombrosis, but they also decrease chances of pregnancy, and pregnancy itself can raise chances of thrombrosis. Menzies admits that chance raising and causation do not always co-incide, but argues that they do for the most part and this is significant.

So Menzies (1996, p. 101) definition of causation goes like this: 'the causal relation is the intrinsic relation that typically holds between two distinct events when one increases the chance of the other event.' This defines the role of the concept of causation in our folk psychology.

The next step is to consider what might occupy this role in the external world. That is, we make an a posteriori identification between this theoretical entity and something else in the world. Menzies suggests that the role occupier might be Fair's conception of causation as energy-momentum

transfer (Fair 1979). To Fair, each case of causation can be reduced to the flow or transference of energy or momentum as it is understood in science between two objects, for example when one billiard ball causes the other to move, kinetic energy is transferred from one to another.

So, a priori, we can say that causation is an intrinsic relation between distinct events where one event raises the probability of the other. A posteriori, we can identify that relation, in our actual world, with energy-momentum tranfer.

In the next section I will move on to Lewis's critique of Menzies (1996). This critique is really Lewis's objections to using the Canberra Plan for causation and does not depend on the details of Menzies' account. However we can keep Menzies' account in mind as an example of a Canberra Plan style analysis.

6.3   Lewis's worries about the Canberra Plan

Lewis gave two reasons for thinking that Menzies' approach, the Canberra Plan, could not work for causation. One of these is the missing relatum objection. I discuss this in chapter two, as one of the possible reasons to be concerned about causation by and of absences. Lewis's other worry is sometimes called the miscellany objection (Liebesman 2011). This is the worry that the Canberra Plan could yield a disjunctive account of causation, and a disjunctive account is a failed account.

Lewis (2004b) argues as follows: Suppose we have two possible worlds, α (actuality) and w (some other possible world). a and b are two events which

occur in w. We need to specify what relation needs to obtain between a and b for us to say that a causes b. One option is that it is whatever fills the role of 'causal relation' in α: that is, the occupier of the causal role in reality. Another option is that it is occupier of the causal relation role in w: another possible world. If the latter is right, then that amounts to saying that the role of 'causal relation' can be occupied by some different relation in each world. This means that we do not *know* that causation itself is an intrinsic relation. In some other world besides our own, it could be another relation which is not intrinsic. It is the nature of the Canberra Plan that we outline and role and then try to locate something which fills it, but it is then possible that different relations could occupy the role in different possible worlds. This means that we do not know the properties of the these different relations in other possible worlds. So, Lewis argues, Menzies argument is supposed to give us the conclusion that causation is an intrinsic relation. But it fails at this job. I will discuss this objection in detail below under the heading of the miscellany objection.

Lewis's worry can be set out as follows:

1. According to the Canberra Plan, we need to first describe the role of causation, and then find out what in reality fills that role. When we have found the role occupier we have our account of causation.

2. The role of causation could be occupied by another relation in other possible world, different from the relation which occupies it in this world.

3. If the role of causation could be filled by some other relation at other

possible worlds, then we would have to list all the role occupiers disjunctively, in order to get our account of causation. We would have a disjunctive analysis.

4. A disjunctive analysis is a failed analysis.

Lewis argues that this problem is insurmountable for accounts like Menzies', but manageable for counterfactual accounts. We can still talk about what would have happened if there had not been an absence, even though the absence is not something which can count as a relatum. To Lewis, this worry is enough to want to avoid using the Canberra Plan for causation.

6.4 Liebesman's response: the folk concept can unify our concept of causation

Liebesman disagrees with Lewis; he argues that the Canberra Plan can be effectively applied to causation. His approach is to deny step (3) of the argument above. According to Liebesman, the miscellany objection is not a serious obstacle. He suggests that when we apply the Canberra Plan to causation we might not be looking for an account of what causation is, or what it is that all these different causal relations have in common that makes them causal. Rather, we might be asking how we came to have a unified concept of causation, despite this apparent disjunctiveness.

Liebesman argues that it is better to interpret Lewis as asking about how we acquired a unified concept of causation, than what unifies causation. That is, about how we came to view causation as one concept, rather than what if

anything makes it a concept which is non-disjunctive. His interpretation of Lewis runs like this:

> First, Lewis observes that instances of causation are widely varied.
>
> From this he concludes that if the Canberra Plan is applied to causation, it will yield something disjunctive and unnatural. ... Second, Lewis notes that we have somehow acquired a fully general concept of causation, applicable to all of its disparate instances, and he claims that it is incumbent on a successful analysis to explain how we acquired such a general concept of causation.
>
> From these two claims, Lewis concludes that applying the Canberra Plan to causation will not meet all of the demands placed on a successful analysis.
>
> (Liebesman 2011, p. 233)

According to Liebesman, we want to know *how we acquired* a unified concept of causation, rather than what it is that all causal relations have in common that makes it possible to give a successful analysis of causation. Liebesman explains this interpretive decision in footnote 7:

> There is a slightly different argument suggested by the last sentence of the paragraph in which Lewis articulates the miscellany objection. He writes 'If causation is, or might be, wildly disjunctive we need to know what unifies the disjunction. For one thing the folk platitudes tell us is that causation is one thing, common

> to the many causal mechanisms' (Lewis 2004b, p. 76). This suggests the following objection. The Canberra Plan will not work for causation because executing it will not help us to explain what it is that unites the common causal mechanisms. This objection is flawed. Discovering the satisfier of the existentially generalized Ramsey sentence and then investigating the nature of that satisfier may help us to understand its underlying nature. (Liebesman 2011, footnote 7)

I agree with Liebesman that some of Lewis's comments make it sound as though he is asking *how* we picked up a unified concept of causation. For example: 'And yet I seem to have picked up a general concept of causation, applicable to all different kinds of causation...' However, Lewis also writes: 'I think conceptual analysis is required to reveal what it is that all the actual and possible varieties of causation have in common.' (Lewis 2004b). This certainly suggests that Lewis is asking how we can successfully analyse causation as one concept, not how we acquired a unified concept.

Liebesman considers it uncharitable to interpret Lewis as asking what "unifies the disjunction". Liebesman writes that it is obvious that the Canberra Plan can tell us what does this. So, Liebesman argues that it makes more sense to interpret Lewis as asking how we acquired a unified concept.

6.5   My response: a folk theory does not make an analysis successful

I disagree with Liebesman that his interpretation of Lewis is more charitable. The Canberra Plan cannot always help us understand what unites different causal mechanisms. Suppose things go like this: we set out our folk theory of causation, turn it into an existentially generalised Ramsey sentence, and then look for the the satisfier of that sentence. But then we discover that there is no one satisfier: rather, there are different satisfiers for different parts of the sentence. Perhaps one relation fills the role of cause when we are talking about straightforward physical causation, but when we are talking about causation between mental states, something else fills the role. Maybe when we are talking about causation between social or economic states of affairs, something else fills the role again.

In this case, the Canberra Plan would not help us see what all these relations had in common. The thing they would have in common would be that they all fit into the cause-shaped role in our folk theory of causation. But we might not find anything in their nature which unified them. Lewis seems to be saying that he thinks there are in fact many different causal relations, and that we cannot be directly acquainted with all of them, and that what "unifies" them is the counterfactual account which we can discover by means of conceptual analysis.

So, there is a perfectly plausible reading of Lewis according to which he is not asking *how* we acquired our unified concept of causation, but what actually "unifies" that concept, if anything. That is, where he is asking what

all the apparent disjuncts have in common which makes them really pone concept.

Liebesman claims that to reach the conclusion that the Canberra Plan will not work for causation, Lewis needs the extra premise:

> EP: Any analysis of causation that yields something disjunctive
> or unnatural will fail to explain how we acquired a general concept
> of causation.
> (Liebesman 2011, p. 233)

Liebesman denies EP. He argues that our folk theory of mind, for example, is as disjunctive as our theory of causation, but Lewis thinks the Canberra Plan is perfectly acceptable as a method for analysing mind. But, according to Liebesman, a folk theory can successfully unite a disjunctive concept. Our folk theory of causation is broad enough to encompass billiard balls striking each other, but also magic spells or causation in a world where everything that happens is caused by God. Liebesman argues that the only thing which "unifies" our concept of causation is our folk theory.

> What explains the ease with which we identify N features and G
> features as causal? The proponent of the Canberra Plan has an
> explanation. Certain features of N and G satisfy the folk theory
> of causation.
> (Liebesman 2011, p. 234)

This depends on what we think it means to unify a disjunction. If we find something which "unifies" a disjunction then that shows that it is not

disjunctive. For example, if we have a property "either has three 60° angles, or has two equal angles and a third angle which add up to 180°, or has three unequal angles which add up to 180°" and we discover that what they have in common is that they are all triangles, we have discovered that our original property was not disjunctive. For almost any property, there has to be a disjunctive way of describing it; the question is whether there is also a unified way of describing it. If we say any property is disjunctive because we could describe it disjunctively then that makes the term 'disjunctive' unhelpful, since it will be applied too widely. So it makes sense to describe properties as unified when we can, and as a disjunctive only when there is no unified way of describing them. It is also the case, though, that we can describe anything in a unified way, if we are prepared to make up a word for some property we would otherwise have to describe disjunctively ("grue"). So whether some property is disjunctive or not depends on the language we are using to describe it.

Another approach to the idea of disjunctiveness is the notion of natural properties. Some properties are natural ones; others less so. So if we want to know if a property is disjunctive or not, we need to first describe it in the most natural way. If this natural description of the property is disjunctive, then the property is disjunctive, and equally if it is unified under its natural description then it is unified. This means that unifying a disjunction would mean finding a natural way of describing it which is not disjunctive.

This is clearly not what Liebesman means, however, since he writes that a folk theory can unify our concept of causation. This makes sense in some ways: we do have a word for causation and a folk theory which tells us what

that word means, so that makes causation unified in the sense that there is some unified way of describing it. However, it's possible that the way to describe in natural terms is still disjunctive. It looks like 'unify a disjunction' to Liebesman just means 'find some non-disjunctive way of describing it', or perhaps some non-disjunctive way that we have an everyday word for and folk theory of. So we are looking for something which is unified relative to everyday language and concepts. But I don't think this is what Lewis means; I think he is after something which is unified relative to our most natural language. These two don't necessarily co-incide: for example, 'pastel coloured' might be disjunctive described in the language of fundamental physics (in terms of different wavelengths of light) but it is perfectly unified relative to everyday language. So, it looks as though Liebesman and Lewis are talking past each other here.

Unity in the sense Lewis seems to be talking about requires more than a folk theory. N features and G features can be both part of a folk theory, but still be disjuncts, if that folk theory can only be described disjunctively in natural terms. A folk theory of causation is only a list of folk platitudes about causation: just a list of things which people think are true of causation. Whether this list of platitudes is joined up with disjunctive connectives, or conjunctive connectives or a mixture of both is contingent on how most people think of causation. It wouldn't be surprising if there were some disjunctive platitudes in the list however: for example, 'c causes e when if c had not happened, e would have happened OR c causes e when c is regularly followed by e'. The folk theory is not a very sophisticated or philosophical thing. Writing out the folk theory of causation would be a dictionary writing or

linguistic anthropology job. The result at the end would tell us a lot about how a group of people think and speak, but nothing about the thing they are thinking or speaking about. The philosophical work begins when we look at the folk theory and try to discover a good reason why those features should be grouped together. If we can't find such a reason then the concept remains disjunctive.

An analogy may help here. I have a shopping list of things to buy: potatoes, parsnips, sprouts, gravy, and stuffing. If we want to know how I know what to buy, we can point to my shopping list. This is Liebesman's point. Want to know how we identify cases as causal? Refer to the folk theory. If we want to know what all these items have in common then we could say 'they're all on the shopping list'. That is true. But it doesn't seem as though it is the kind of answer the questioner wanted. If we wanted an explanation for why I am buying those items, saying 'because they are on the shopping list' would be woefully inadequate. We need a more substantive answer: because I am making roast dinner. It seems as though this would really be a better answer to the question of how I know what to buy as well; after all, the shopping list is only an intermediate stage. So similarly, if we want to know how we identify all these cases as causal, saying 'they are all causal according to the folk theory' doesn't seem like the answer we really look like the answer we want. We want something more substantive. We want to know how we came to have that folk theory of that concept.

To conclude this section, Lewis does not think the Canberra Plan can be applied to causation, in part because of what Liebesman dubs the miscellany objection. Our concept of causation seems disjunctive, which makes

us wonder how we could have acquired a unified concept of causation, but also what it is which all cases of causation have in common which makes analysing causation as one concept a good approach.

Liebesman argues that Lewis is asking *how* we acquired a unified concept of causation, while I think Lewis is best interpreted as asking *what* it is which makes the concept of causation non-disjunctive. A folk theory does not give us an answer. It can only tell us that we group a certain set of features together, not why.

6.6 Another way to deal with the miscellany objection

In my view, the way to deal with the miscellany objection is to deny (2). That is, we need to deny that the role of causation could be filled by another relation at another possible world. I will argue that the worry that causation might be filled by something else at another possible world is not something we need to be worried about unless that other possible world is a genuine metaphysical possibility. I will argue that the examples which Lewis gives are not up to the job: two of them metaphysically impossible due to logical flaws, and there is nothing to suggest that the other is metaphysically possible either. I will then argue that fully imagining a world where the laws of nature are different demands more than stipulating what the world should look like. Finally, I will set out the steps of my argument that we can deny that the role of causation is filled by something else at a another possible world.

First, let us look at the example Lewis gives of other possible worlds where the role of causation is filled by something else from what fills it at this world. It is not completely clear what Lewis's argument is for thinking that the causal role could possibly have a different occupant. On page 280 he says that presumably the occupier of the role is a 'contingent matter'. Perhaps the implied thought is that we have to look empirically at the world around us to find the role occupier, and things which we find out empirically are usually contingent truths. A response to this is to suggest that the identity of the role occupier is a necessary truth discovered a posteriori; I will argue for this view below. Later, on page 284, Lewis describes a selection of other relations which he thinks could have been role occupiers. It looks as though the general thought here is that we can imagine worlds where something else occupies the role of causation, and therefore we cannot rule them out.

However, I will argue that imagining a possible world, in a way which shows anything interesting philosophically, is different from just describing a world with a collection of stipulated features. Before positing that world proves anything for our thought experiment, we have to show that those features are consistent with each other. This would involve imagining the world in much more detail than we are normally in a position to do. So, Lewis's examples of what might have filled the role of causation in another possible world do not show anything about whether something else could have occupied the role of causation.

Let us look at the details of Lewis's examples. He gives several (Lewis 2004b):

- Aristotle World, where the laws of physics follow Aristotelian dynamics: 'the occupant of the biff-role in such a world might be a kind of 'force'... which is proportional to velocity rather than acceleration.'

- Impetus World. Instead of transfer of energy/ momentum, we have transfer of impetus, where impetus fades away spontaneously instead of being conserved.

- Occasional World. In this world occasionalism is true; every causal relationship is mediated by God's willing the effect to happen.

Let us begin with Aristotle World. There is no space to give a detailed account of Aristotelian physics here, but I will give a brief sketch. According to Aristotle, each movement of an object is contingent and must be caused. The movements may either be against the nature of the object or according to its nature. Each object has a natural place in the universe depending on what it is made of. It is in the nature of light bodies to move up and heavy bodies to move down.

Aristotelian physics, however, has some inconsistencies. For example, according to Aristotle, heavy bodies fall faster than lighter bodies. But Galileo used a thought experiment to show that this cannot be right. If we attach a heavy object to a light object the result should fall faster because of the heavy object. But if we attach a light object to a heavy object the result should move more slowly due to the light object. So, it looks as though a heavy object and a light object joined together should both move faster and more slowly. This is logically impossible, showing that Aristotelian physics is inconsistent, and thus not a metaphysical possibility.

Next, let's consider Impetus World. The conservation laws are essential to our world's working as it does. A world where impetus is something which can be transferred but which can also fade away would be utterly different. An important question then would be whether such a world would be fading, as the impetus in it gradually trickled away, or whether there are sources which spontaneously create impetus in the same way that other objects spontaneously lose it. We could also ask questions about when and at what rate objects lose impetus, and whether the spontaneous gaining and losing of impetus should also count as causal. Without some details for how this world would work, all this is pure fantasy. We could stipulate anything we want, but without a detailed understanding of how the physics in this world would work, we have failed to imagine it. And therefore, we cannot see whether anything like causation could occur there.

Finally (for now) let us take a look at Occasional World. In this world, God has to mediate every cause and effect. To fully imagine this world we have to say how it is, exactly, that God mediates these cases of causation. Perhaps it functions like comment moderation on a discussion forum: cases of cause and effect queue up, and God allows them as he sees fit. Yet this still doesn't tell us how some effect joins the queue of effects waiting to be allowed by God, and how it is that God allows each effect. In order to allow or not allow each effect God has to intervene on the world in some way, and that intervention sounds, in itself, causal. So then we need to ask who allows the instances of second order causation which consist in God allowing first order causation to go ahead. We need to decide if there is a meta-God, allowing God to have a causal role, or whether perhaps there is a special kind

of causation which God uses to allow effects to occur, which does not itself require divine permission. Again, the ludicrous nature of these questions show that without some detail theologico-physical understanding of what this world would be like, we have not fully imagined it, and therefore just stipulating some features which it might have doesn't prove anything.

So, Lewis gives three examples of supposedly-possible worlds which are meant to show that something other than biff could ground causation. But out of these three, two of them – Aristotle World and Occasional World – have logical problems, and are therefore not genuine metaphysical possibilities. Impetus World has a lot of unanswered questions, and we have no reason to think that it is any more metaphysically possible than the other two suggestions, especially with only the vaguest hints at what it might be like.

At this point, one response would be to argue that surely we can imagine a world where the laws of physics are otherwise and something other than biff grounds causation. We could stipulate that there is a world, where there are causal relations obtain, and where they are grounded by something other than biff, and where there are no logical inconsistencies of the sort that would make the world metaphysically impossible. These stipulations do not appear mutually contradictory on the face of it. However, I think that merely stipulating that a possible world like this exists is not the same as showing that it is a genuine metaphysical possibility. In order to show that it is a metaphysical possibility, I think a world needs to be fully imagined. Imagining a possible world in the detail required to show that it is a metaphysical possibility would be demanding and complex.

Here is an argument that imagining the laws of nature to be otherwise is

more complex than simply stipulating that there is a possible world where they are otherwise. Take the fact that the universe is expanding and all the stars are moving away from us. What is it to imagine that the stars are *not* moving away? With one way of imagining things being otherwise, it takes no effort at all. It doesn't look as though the stars are moving away, in the same sense of "looks like it is moving away" that if I stand on the curb and watch a car pull away it looks as though the car is moving away from me. We don't look up at the night sky and say 'look at the stars, all flying away from us' unless we are nerds. The stars look like they are rising in the east and setting in the west, and at a quick glance they don't look like they are moving at all. But, on another way of imagining things otherwise, if the stars were not moving away from us the night sky would not be dark. Without the stars moving away, more of their light would reach us, so that the whole night sky would shine. Thought of this way, when we look up at the sky, it does look as though the universe is expanding; we just need the right idea of what the universe's expanding looks like in order to be able to see that. On a third view, it might not even make any sense to ask what the universe would look like if it wasn't expanding, since we might not be able to exist in a non-expanding universe. On this view, there is no 'how it would look for the universe to not be expanding' because for anything to be looking, the universe must be expanding.

We need to avoid confusing 'what is it like?' in the sense of 'what state of affairs obtains?' and 'what is it like?' as in 'how does it appear?' There are three questions we can ask:

1. How do things appear?

2. What state of affairs must obtain in order for things to appear this way?

3. What state of affairs obtains?

Question 1 is strictly about appearance, so of the night sky we could only say 'dark, with points of light', not 'like the stars are standing still'. It's a truism that whatever state of affairs needs to obtain in order for things to appear the way they actually appear, is the state of affairs that appears. (Thus there is a sense in which things are always as they appear; we just don't always know what state of affairs must obtain in order for them to appear that way.) In the case of the night sky, if we think that it looks as though the stars are standing still, we do not understand what the sky would look like if the stars were really standing still. We are talking as though a state of affairs which would ensure the stars' appearing to stand still would be the stars' actually standing still. But this is false. If the stars were really standing still, everything would appear otherwise than how it does, or perhaps there would be no one for it to appear to.

So, imagining the laws of nature to be different is a more complicated task than it first appears to be. We can easily imagine a world where (a) the laws of nature are otherwise and (b) the laws of nature's being that way is enough for the world to appear the way it does already. That is, we can imagine it as long as we do not have to provide any details for how (b) works out. Without the details, there is not much to this imagining. It amounts to: 'imagine a world which is different from this one, but appears indistinguish-

able'. Without the details, we have not shown that this supposedly-possible world is a genuine metaphysical possibility. If we did want to provide details of this world, showing how it really works, we would need to know more about this other world that we do about the actual one. After all, we don't yet have a full story for our own world of how the way things are brings about the way they appear. (Does this mean that we cannot show that our world is a genuine metaphysical possibility yet? Yes. Luckily, we have other evidence that our world is metaphysically possible: it is actual.)

To conclude this section, here is my argument that we cannot imagine a metaphysically possible world where causation is grounded by something other than what grounds it at this world.

1. To imagine a world with physics significantly different from our own, we would have to also imagine all the consequences of that physics being different.

2. Without constructing a hypothetical physics of that other possible world, we cannot imagine all the consequences of that world's physics.

3. Part of the consequences of a world's having very different physics will be how that world appears.

4. So, without constructing a hypothetical physics of that other possible world, we cannot imagine how that world would appear.

5. So, we cannot imagine just any possible physics with just any possible way for that world to appear.

6. We cannot imagine a world with significantly different physics to our own. (Or at least, not without constructing a hypothetical physics for that world, which we have not done.)

7. A world where the role of causation was occupied by some other relation from what occupies it in this world would have to have physics significantly different from our own.

8. We cannot imagine a world where the role of causation is occupied by something other than what occupies it at this world.

To conclude this section, we do not need to worry about the miscellany objection, because we can deny that the causal relation could be filled by something else at some other possible world. This argument will come into play later in chapter 11.

6.7 Conclusion

Although Lewis was concerned about using the Canberra Plan as an approach to analysing causation, I have argued that we can use the Canberra Plan or something like it without being concerned about ending up with a disjunctive analysis. In some ways, the approach I will end up suggesting in this thesis resembles the Canberra Plan. I will suggest that there are two steps to analysing causation, and that there is a grounding relation between the two parts of the analysis. (In the Canberra Plan the role occupier presumably grounds the causal role.)

Any account with this two tier structure will have to deal with miscellany objection style worries. Very generally, if there are two parts to the analysis of causation: the more fundamental causation$_1$ and the less fundamental causation$_2$, and causation$_1$ is supposed to ground causation$_2$, then we will need to answer the question of whether something besides causation$_1$ could have grounded causation$_2$. Since in chapter eleven I suggest a two tier account of causation, my arguments in the section above will be important to my conclusion.

The account which I endorse will be different from the Canberra Plan. The Canberra Plan still does not explain our conflicting intuitions about causation, which I discuss in more detail in chapter ten. The Canberra Plan really only suggests one concept of causation, which we partly understand through specifying a role and partly through identifying a role occupier. I will suggest a two-tier account of causation, where we have two concepts of causation. I will defend this view in chapter eleven.

Part III

Defending reductive analysis

Chapter 7

Woodward's non-reductive account

7.1 Introduction

In this chapter I will give an outline of Woodward's non-reductive analysis
of causation. I will look at the case for providing a non-reductive analysis
of causation, and Woodward's argument that this is every bit as valuable as
reductive analysis would be, and more achievable. I will then discuss some
problems with these arguments and with this approach. Highly developed
accounts such as Woodward's show us that non-reductive accounts must not
be dismissed out of hand, so I suggest that if we are to favour a reduc-
tive account, this needs to be argued for. In this chapter I will argue that
Woodward's claim that a non-reductive analysis can be just as good as a
reductive analysis has some problems with it. In the following two chapters
I will discuss two possible reasons for preferring a reductive analysis to a

non-reductive analysis.

First, in section 7.2 I will briefly explain some background to Woodward's account: manipulationist accounts of causation. Next, in 7.3 I will describe Woodward's aims in analysing causation; that is, what he considers a good analysis of causation. I will then give an outline of Woodward's account in 7.4. After that, in section 7.5, I will briefly discuss the role of causal modelling in the analysis of causation. Finally, in 7.6 I will discuss Woodward's defence of his approach of giving a non-reductive account.

7.2   Manipulationist accounts

Historically, one approach to causation has been manipulationism. According to this approach an event is a cause if we can use it to manipulate some other event. For example, turning on the oven allows us to control the cake's baking.

One advantage of this approach which Collingwood (1998) points out is that it allows us to account for differing perspectives on what a cause is. Take the case of a speeding driver who skids on the winding icy road and ends up in the ditch. What is the cause of the crash? A roadbuilder might say the it was the sharpness of the bends in the road. The driver might admit that the cause was her going too fast. Both factors seem causal, but to have two causes seems like overdetermination. Collingwood writes that the point is that we think of the cause as the bit we could have used to manipulate the outcome. So if a roadbuilder wants to prevent car crashes, she needs to think about the sharpness of bends in the road. But if a driver wants to

prevent crashes, she needs to think about speed. If what we see as the cause depends on what we could manipulate, then this explains why sometimes there are different causes from different points of view. Von Wright (1971) takes a similar view.

However, there are problems with the manipulationist approach to causation. It is either unacceptably circular, or unacceptably anthropocentric. On the one hand we can try to break down the concept of manipulation. To manipulate something means to bring about a change in it; it is a causal concept. To define causation in terms of a causal concept in this way is circular. To avoid this circularity, we could take our concept of manipulation as primitive. But the concept of manipulation involves human beings, or at least agents of some sort; things can be manipulated by animals, and the idea of them being manipulated by angels is perfectly coherent, but not the idea of a thing's being manipulated by a rock or a magnetic field. But surely we want to say that events involving no agents can be causal.

Woodward's account can be seen against the background of these unsatisfying manipulationist accounts. The appealing idea behind manipulationist accounts is that the cause is the event, the handle if you like, which we can use to manipulate the effect. This sounds right, and it also does not sound immediately circular. Manipulationist attempts to unpack this idea have been unsatisfying however. Woodward's approach is to accept that the idea of an intervention is causal, and thus that his account is circular, but to develop a complex set of criteria for when a relationship between events counts as causal, with the aim of providing an illuminating non-reductive account. I will discuss these issues of non-reductive analysis and illuminating

(or not) circularity futher below. In the next section I will move to discussing Woodward's account.

7.3   Woodward's aims in analysing causation

In this section I will outline Woodward's aims in analysing causation and try to show what motivates him to analyse causation in the way that he does.

Woodward (2003) lists his aims in producing an account of causal explanation, which I will summarise here:

1. The theory should describe the uses of causal explanations in everyday life and in science, and should account for why an explanation works as an explanation.

2. If there turns out to be more than one kind of causal explanation, the theory should explain why they are all different kinds of the same thing.

3. The theory should be able to distinguish explanations from descriptions, and better explanations from worse ones.

4. The theory should be able to handle objections which have been faced by previous philosophical theories of causation, and make sense of their successes and failures.

5. The theory should explain how we know what we know. If we think a certain group of people who are using causal language and making causal inferences correctly, and we think that making causal inferences and using causal language means knowing and understanding certain

things, it must be the case that those people do know and understand those things.

With regard to point (5) above, Woodward is saying that we cannot attribute causal understanding to children and animals if we define cause in terms of something children and animals don't understand. And the way we draw causal inferences in science is also something the theory must explain.

The target of Woodward's analysis, then, is very much the concept of causation, rather than the word. Woodward argues that any creature that aims to manipulate the world around it, must have some concept of cause; this includes animals and very young children. Equally, when a scientist draws some causal inferences from data, she may not explicitly use the word 'cause'; nevertheless the concept of cause can be present in practices with 'a substantial non-verbal element' (Woodward 2003). Woodward is aiming to capture this concept.

Many of Woodward's criticisms of the mark transference theory due to Salmon (1984) are also applicable to Dowe's conserved quantity theory. For example, Woodward is unimpressed by a theory which cannot allow for action at a distance (for example, of gravity), omission and double prevention, as well as psychological and economic causes. Woodward (2003, p. 353) also points out that according to Salmon's theory (which resembles Dowe's in many important ways) the correct causal explanation for gases in a container expanding as they heat up would involve tracing the causal processes of each molecule of gas. The usual scientific explanation does nothing of the sort. So, Woodward argues, Salmon's account does not really account for

why scientific explanation work the way they do; according to his account scientific explanation ought to be more reductive than it in fact is.

I believe that there is an assumption here which can be challenged: we are assuming that in order to pick out the cases where a concept applies, in order to use it, a person must also understand what grounds that concept. I will argue against this assumption is more detail, and explore some implications of challenging this aspect of Woodward's approach, below in the final section of this chapter.

Finally, Woodward takes the view that an account of causation should ideally explain why we care about causation, as well as what causation is. So for Woodward, transference theories such as Salmon's are unsatisfying, because they do not account for why we care about quantities of transferred energy or transferred marks. On the other hand, a theory which defines causation in terms of manipulation also explains why we should care about it, since we want to be able to manipulate the world around us. Again, I believe this can be challenged and will argue against it below.

7.4   Woodward's account of causation

In this section I will summarise the core of Woodward's account of causation. I will begin by giving a rough explanation of how Woodward's account goes, and will then give his more precise definitions of intervention variables, interventions and invariance.

So Woodward's account is supposed to tell us when causation is occurring between two variables: X and Y. In order to explain Woodward's ac-

count, however, we also need to think in terms of I, the intervention variable. Changes in X may be followed by changes in Y, but the account is meant to tell us in the change in Y is *caused* by the change in X. For example, suppose Y is the health of an indoor plant: the variable Y might have the value 'healthy' or 'unhealthy'. X is the light level around the plant: it could be dim or bright. Maybe we have observed that the plant looks unhealthy in winter when the light coming through the window is dim, but healthier in summer when the sun shines. How would we find out if the dimness is causing the plant's sickness? It's obvious that we could experiment by altering the amount of light. This is where I comes in; let's say I is the curtains being open or shut. We could change the value of the intervention variable I by closing the curtains to set the value of X to dim, and then look to see whether the relationship we had observed between X and Y still held true. In Woodward's terms, we are finding out if the generalisation, that X is followed by Y is invariant under the intervention I.

These general ideas are the bare bones of a scientific experiment. But according to Woodward they also make up the basic idea of what causation is. The relationship between X and Y is causal if it stays invariant under certain interventions on X using I. There are certain caveats: I must not change Y in any way that does not go through X for example. To summarise Woodward's view, X causes Y when there is a relationship between X and Y which is invariant under certain interventions on X using I. Obviously to make sense of this we need to understand what Woodward means by 'intervention'.

It is important to note that the concept of an intervention is not a human-

centred one. So in the example above, it is a matter of my closing or opening the curtains. But the intervention is not defined as something a person does. Rather, it is a third variable which has a certain relationship to X and Y. One way of thinking about it is that X and Y form a self-contained system. To understand anything about how that system works on a counterfactual and causal level, we have to change something from outside. The intervention variable may or may not be something that is done by a human being, but it always breaks any causal links between the normal cause of X and X itself, thus changing the system in a surgically specific way.

At this point it will be useful to see Woodward's exact definitions of interventions. The exact wording is important, and it is also useful to see exactly where and how the notion of cause is used in defining an intervention.

He begins with a type-level definition of an intervention variable, and then goes on to describe an intervention, at token-level.

I is an intervention variable for X when:

(IV)

I1. I causes X.

I2. I acts as a switch for all the other variables that cause X. That is, certain values of I are such that when I attains those values, X ceases to depend on the values of other variables that cause X and instead depends only on the value taken by I.

I3. Any directed path from I to Y goes through X. That is, I does not directly cause Y and is not a cause of any causes of Y that are distinct from X except, of course, for those causes of Y,

if any, that are built into the $I - X - Y$ connection itself; that is, except for (a) any causes of Y that are effects of X (ie variables which are causally between X and Y) and (b) any causes of Y that are between I and X and have no effect on Y independently of X.

I4. I is (statistically) independent of any variable Z that causes Y and that is on a directed path that does not go through X.

Given the notion of an intervention variable, an *intervention* may be defined as follows:

(IN) I's assuming some value $I = z_i$, is an intervention on X with respect to Y if and only if I is an intervention variable for X with respect to Y and $I = z_i$ is an actual cause of the value taken by X.

(Woodward 2003, p. 98)

The parts of the definition of an intervention which involve the concept of cause, phrased more roughly, are as follows:

1. When the intervention occurs, X stops being caused on whatever it used to be caused by, and only depends on the intervention variable.

2. The intervention variable must not cause Y directly, or cause anything which causes Y except for X, unless it is part of the chain of causes between I through X to Y.

3. *I* must be independent of any variable that causes Y in a way that does not go through X.

One thing immediately worth noting is that the causal link between X and Y is not being defined in terms of itself. Instead, it is being defined in terms of the presence or absence of other causal links. This makes the problem of circularity seem less pressing, although overall this is still intended to be an account of causation in general, not a characterisation of a particular causal link.

The other key notion which Woodward uses to define causation is the idea of invariance. A causal relationship, to Woodward, is a relationship which is invariant under certain interventions. This is perhaps a more generalised version of the idea of a law of nature. A law of nature is meant to be invariant under all interventions: to hold true universally. An invariant generalisation is similar but not universal; it only holds true in some specific cases.

To sum up, for Woodward, causation is best thought of in terms of one variable causing another to take a different value, or in terms of a relationship between two variables which is invariant under certain interventions. So the concept of causation rests on the idea of an intervention. This cannot be understood in purely non-causal terms, but nevertheless gives us a picture of how, knowing some causal and non-causal facts, we can discover new ones.

7.5   Non-reductionism in causal modelling approaches to causation

Woodward's approach is inspired by an approach to understanding causation in the sciences, especially the social sciences, known as causal modelling. Causal modelling is an approach to finding causal relationships between variables used in various sciences, especially medicine and the social sciences. It is used to identify causal relationships between variables in large amounts of data, for example, to identify a causal relationship between patients receiving a certain treatment and a certain outcome for their health. Causal modelling is a method for understanding the causal structure of a situation, in cases where the web of causally connected events is more complex than c causes e.

In recent years some philosophers have looked to causal modelling for suggestions about how we ought to analyse what causation is. Causal modelling can allow us to specify a particular kind of counterfactual, "one whose antecedent specifies values for arbitrarily many variables" (Paul and Hall 2013, p. 18), and to give truth conditions for these counterfactuals. The idea is that the counterfactuals thus specified (interventionist counterfactuals) can be used to analyse causation. Interventionist accounts of causation, are then a species of counterfactual accounts, but using this specific kind of counterfactual.

Causal modelling is an approach to understanding webs of causal relationships, and does not by itself commit us to any particular account of what causation is. Rather, certain views lead philosophers to use causal modelling in their analysis of causation. An examples of one such view is the belief that

our account of what causation is should be closely linked to our account of how we come to find out about specific causal links.

Another example is the concern that our account of causation should make sense of the different kinds of causal link, rather than just give an account of the one relation, causation. One way of looking at it is that interventionism is an account of what causation is which explains why, in science, causal modelling works. Below, however, I will argue that this is not the job that an analysis of causation needs to do.

A causal model consists in variables, a range of possible values for each variable, which variable depends on which, and structural equations that specify which variables depend on which, and in what way. So in the classic case of Suzie throwing a rock at a window, the variables are whether or not Suzie throws and whether the window breaks. The possible values for the first variable are "Suzie throws" and "Suzie doesn't throw" and for the second, "the window breaks" and the "window doesn't break". Even with this simple example we can see how the variables could take other values rather such as "the window is cracked". In the simplest cases, variables may only have two values, 0 and 1, but in other cases they may have many different possible values. The values of some variables depend on the values of other variables and this information can be captured in structural equations. Causal models can be represented using directed graphs. Each node corresponds to a variable.

7.6 Woodward's non-reductionism

Most accounts of causation are intended to be reductive. One way of understanding this is to consider the cluster of related concepts: 'cause', 'law of nature', 'counterfactual dependence', 'power' and 'disposition'. A reductive account of any of these concepts would try to either define the concept in terms of concepts outside that cluster (for example, mark transference, correlation or constant conjunction) or define the concept in terms of some concept within the cluster which could, in turn, be analysed by a concept outside it.

Examples include Dowe's account, where the difference between causal and non-causal processes is based on whether certain empirical physical facts are true about them (Dowe 2000). In Ducasse's account, what counts as a cause is defined by its being the only event spatiotemporally contiguous with the effect (Ducasse 1969). Even Lewis's counterfactual account has this reductive aim: even though he uses the concept of counterfactual dependence, this gets reductively defined elsewhere in terms of possible worlds (Lewis 1973b).

In the case of fundamental concepts, we can define them non-reductively, in terms of concepts within the cluster. Fundamental concepts are not reducible to anything else. Some concepts, such as identity and duplication seem so inherently bound up with each other that we cannot define either of them without reference to the other. In this case the non-reductive analysis helps us understand how the concepts in the cluster relate to each other. The account can be described as an illuminating non-reductive account.

Woodward's view is not that the concept of causation is fundamental, but that a satisfactory reductive analysis will be very difficult to find, if not impossible, and so a non-reductive analysis is the best approach. One obvious criticism of Woodward's argument is that it is unilluminatingly circular.

Woodward has a number of counter-arguments against the view that his account is unhelpfully circular. One is that if non-reductive theories were trivial, they would not be able to conflict with reductive theories, and yet they do. A view of causation, for example, which says that causation is production (and says of production that it is the same as causation) does not really conflict with any other theory because there is nothing substantive to it; it's really just that two words in our language are synonyms.

Woodward's other point is that even when we are defining one concept in the cluster with other concepts in the cluster, there is still more than one way to do that. For example, Woodward points out that many philosophers think that there is a link between causation and counterfactuals, but Woodward (2003, p. 21) suggests that in some of the previous literature we have been linking causation with the wrong counterfactuals. For Woodward, questions about how the link between causation and counterfactuals works, and which specific counterfactuals are involved, are much more interesting and important than the question of whether we can achieve an ultimately reductive account through analysing causation in counterfactual terms.

If non-reductive analysis is the positing of a grounding relation without taking a stand on the direction of that relation, then non-reductive analysis is not trivial. But it does mean saying less than a reductive analysis would. If there is a grounding relation then it has a direction, since grounding is an

asymmetric relation, so an account which is able to say what the direction of the grounding is as well as that there is a grounding relation, is a more comprehensive account. On this view, non-reductive accounts are not trivial, but they are limited. It seems likely, though, that this is not what Woodward has in mind when he talks about a non-reductive accounts; he writes as though he sees non-reductive accounts as ultimately going round in a circle, so that there is a direction of analysis. The account is worthwhile if it illuminates the concepts involved.

With this in mind, we can ask why Woodward offers a non-reductive account. The answer to this is in the criteria he sets out for a good account of causation. He believes that a reductive account will not be able to meet these aims. So we should discuss whether these criteria are good ones, and whether they are worth the cost of providing an analysis which is non-reductive.

Here are Woodward's criteria again:

1. The theory should describe the uses of causal explanations in everyday life and in science, and should account for why an explanation works as an explanation.

2. If there turns out to be more than one kind of causal explanation, the theory should explain why they are all different kinds of the same thing.

3. The theory should be able to distinguish explanations from descriptions, and better explanations from worse ones.

4. The theory should be able to handle objections which have been faced by previous philosophical theories of causation, and make sense of their

successes and failures.

5. The theory should explain how we know what we know. If we think a certain group of people who are using causal language and making causal inferences correctly, and we think that making causal inferences and using causal language means knowing and understanding certain things, it must be the case that those people do know and understand those things.

A principle very much like (1) has been contested by Strevens in his book 'Depth'. Strevens (2008) argues that it is a mistake to try to fold the structure of causal explanation into the metaphysics of what causation is[1]. The thought is that our analysis of what causation is need not reflect the structure of our causal explanations. Strevens is much more interested in the structure of our explanations than the metaphysics of causation. This is also true of Woodward, but where Strevens leaves the details of the metaphysics to others, Woodward argues that they should follow the structure of causal explanations. The same thought goes for (2) and (3).

None of these points, however, really explain why Woodward needs to provide a non-reductive analysis. These criteria seem like things which most accounts of causation, reductive or not, would aim to do. (Especially point (4) is something that every philosophical account of anything tries to do: build on previous accounts and explain their shortcomings.) Clearly Woodward

---

[1]Note that the view Strevens describes in 'Depth' is a little different from the view he espouses in his more recent paper 'Causality Re-unified' which I discuss in detail in chapter 11. In his book, causal difference making is not a form of causation but of causal explanation.

feels he has a better chance of meeting these criteria with a non-reductive account, but he does not say why a reductive account will do this job better.

Woodward argues that a non-reductive account of causation is the only account which could meet his criteria for a good account of causation. The thought is causal modelling works because it gets the structure of causal relations right. Many accounts of causation, historically, have taken for granted the structure '*c* causes *e*'. In the same vein, Hitchcock (2007) argues that it is better to look to the more complex structures described using the causal modelling approach: a web of cause and effect, where each causal relatum can be understood as a variable which can take one of two values or more, the relationship between these variables can be described using equations. Hitchcock argues that since we are able to successfully describe causal relationships in science using causal modelling, the '*c* causes *e*' structure is overly simplistic and our account of causation should be based around the structure provided by causal modelling, as with Woodward's account above.

However, a number of writers have argued that we can base the structure of causation on causal modelling whilst getting the metaphysics from somewhere else. Handfield et al. (2008) argue that we should get our metaphysics of causation from process accounts such as Dowe's. I will return to their view in chapter 11. Glynn (2013) argues that causal modelling analyses can be combined with Lewis's possible worlds semantics of causation. In either case, we would not need to give up on providing a reductive account of causation.

I have not yet said, however, why we should want a reductive analysis. I have said that Woodward does not supply adequate motivation for giving a non-reductive account, but not what would be so bad about an account's

being non-reductive, or looking at it from other point of view, what is so good about reduction. I will address this in the next two chapters. First, I will suggest that one motivation for reductive analysis of causation might be ideological, but I will argue that this is not a good reason. Secondly, I will argue that empiricism does provide us with a good motivation for seeking a reductive analysis.

Chapter 8

Ideological parsimony

8.1 Introduction

In this chapter I want to discuss one possible reason for wanting to provide
a reductive analysis of causation: ideological parsimony. My view is that
while the pursuit of ideological parsimony does seem like a good reason at
first, closer inspection shows that it does not provide the motivation we are
looking for.

I will start off by explaining briefly what ideological parsimony is and why
at first it seems like it should provide a motivation for reductive analysis. I
will then go on to examine two problems with it. Firstly, it is hard to assess
when an hypothesis is more or less parsimonious and this makes it difficult to
compare two hypotheses and say which one is more parsimonious. Secondly,
even in cases where we can tell which hypothesis is more parsimonious, in
the philosophical context, we lack a good justification for why we should
favour that hypothesis. At best, we can justify a preference for parsimony in

science, but not philosophy.

Parsimony is akin to simplicity. Often simplicity is used to describe some vaguer concept, while parsimony is meant to be something more precise. In this chapter I will use the terms interchangeably, and treat complexity as their opposite.

Within parsimony, we have ontological parsimony, where we try to avoid positing too many existent things, and ideological parsimony, where we try to avoid positing too many ways of being. An ideological primitive is a fundamental property or a relation which cannot be reduced to other properties or relations; we use our ideology to describe what the world is like. This distinction between ideology and ontology is due to Quine (1951). An ideologically parsimonious theory, then, will be a theory which gives us a description of what the world is like using the fewest possible primitives.

Our metaphysical theory, overall, will include both ideology and ontology. If causation is fundamental, then it will be in our theory as a part of our ideology, since it is generally taken to be a relation. So it seems a natural thought that, if we could get rid of causation from our ideology, that would be good for parsimony. On the other hand, we want to go on talking about causation meaningfully. So we want to eliminate causation from our ideology in one sense, while retaining it in another.

A realist might approach this differently. According to a realist, properties and relations are part of the ontology of a theory too. Most of this chapter will be from a nominalist perspective, but all the points translate for the realist. Where a person stands on realism vs nominalism about relations will not change whether or not they can accept my conclusions.

With a reductive analysis of causation, we will be able to reduce it too, some other primitive or primitives which we would have had in our ideology anyway. For example, we might take a Lewisian approach and reduce causation to counterfactuals, and then reduce counterfactuals to similarity relations between possible worlds (which are just like our world but spatiotemporally disconnected). Then Lewis is just left with similarity and spatiotemporal connectedness in his ideology, but we can still talk about causation and make sense, with the reductive analysis in place. We can see how this strategy could lead to greater ideological parsimony in our overall metaphysical theory.

A distinction between types of parsimony is between quantitative and qualitative parsimony. Lewis (1986) argues in his book 'On the Plurality of Worlds' that possible worlds realism can be defended on the grounds that while it involves many entities (poor in terms of quantitative parsimony) it posits few kinds of entities, since all possible worlds are the same kind of thing (good in terms of qualitative parsimony). Lewis argued that qualitative parsimony is the important one. Nolan (1997) argues that quantitative parsimony is also an important theoretical virtue. Whilst I do not plan to discuss these issues in detail here, it is interesting to note the different varieties of parsimony and the lack of agreement among philosophers about how to rank their importance as theoretical virtues.

So that is one possible reason why we would want a reductive analysis: it aids ideological parsimony. However, there are a couple of questions we need to address: can we really measure ideological parsimony? and if we can, what reason is there to think it makes a theory more likely to be true?

8.2   Measuring ideological parsimony

Firstly, let us discuss the problem of measuring ideological parsimony. There are two ways to go about this. One is a holistic, response dependent approach. The other approach is to coin a term for a unit with which we can measure the parsimony of an hypothesis. For this to work, a theory must be made up of components, and we can take the ideological side of our theory to be made of ideological primitives. Then, measuring the parsimony of the theory will be a matter of counting the number of ideological primitives.

The holistic, response dependent approach means that when we are weighing up two theories to discover which is the more parsimonious, we just consider them both and make a judgement about which one seems simpler. For example, we might weigh up Lewis's theory against a counterfactual analysis of causation which leave counterfactuals unanalysed, and say 'Lewis's is simpler'. How do we know? It just seems simpler. This approach to parsimony will give us problems for both measurement and justification, but I will focus on the former for now. We could easily have a situation where two people disagreed about which theory was the simpler, and there would be no way to reconcile their disagreement. It would make simplicity a matter of taste, and measuring simplicity would be as impossible as measuring the attractiveness of interior decoration. We cannot use ideological parsimony as a motivation for adopting a particular kind of analysis if we do not know when we have achieved ideological parsimony.

This leaves us with the theoretical components approach. This seems initially a lot more promising. There is already an obvious theoretical com-

ponent to count: the ideological primitive. And it seems fairly easy to tell what one ideological primitive is, as well. We need only look at all our property and relation terms and see which reduce to or can be identified with others. Whatever we have at the end is our list of primitives.

However, this approach comes with an exchange rate problem. Here's the issue. The theory which we considering, and trying to make as parsimonious as possible, is our overall metaphysical theory. That needs to have an ontological side as well as an ideological side. Even if we are ontological nihilists, and think that there are no things, only ways for things to be (and that sounds incoherent) we will still need to have a part of our theory that talks about ontology, even if it simply says 'nothing exists'. So the theoretical component for our overall metaphysical theory cannot be the ideological primitive, because the theory as a whole is not all about ideology. Quine (1963) notes that the derivative concepts of one theory may be the basic concepts of another, and vice versa.

Now, the obvious response here would be to say 'ah, well if the ontological and ideological sides of the theory have different theoretical components then they are different theories.' But this will not do, because it is clearly possible to shift the burden between ideological and ontological areas of a theory. In other words, we can choose how much of our metaphysics to work into the ontology side of our theory and how much to work into the ideological side. When we are picking out what is needed to describe the world at a fundamental level, we can decide how much of that is going to be lists of what is and how much is ways for it to be.

A classic example of a theory which is ideologically heavy but ontologi-

cally light would be the ontological nihilism suggested by O'Leary-Hawthorne et al. (1995). On this view, there are no things but only ways for things to be. An example of a theory which is arguably ontologically heavy but ideologically light is Lewis's possible world realism (Lewis 1986) . That is, he posits many possible worlds. But because he does that, he does not then have to posit possible being and actual being. To be possible is just to exist at some other world. His defence against the charge of ontological unparsimoniousness is that his theory is qualitatively ontologically parsimonious (because all the possible worlds are one kind of thing, even though each is a different thing), and it is also ideologically parsimonious, because he has analysed away modality. Our theory of modality, whether it is possible worlds realism or some actualist story, will have ideological and ontological components.

The notion that our ontology only contains particulars and properties, whereas relations belong in our ideology is a nominalist perspective. A realist about universals would count properties and relations, including the causal relation, as objects in our ontology, along with particulars. So we might be tempted to wonder if we can solve the exchange rate problem by being realists. However, I think in this case we have just shifted the theoretical burden again towards the ontological side. We would still need an ideology in this theory. For example, we would need instantiation to explain how we connect causation and the particulars which it relates. Treating instantiation as a real object in our ontology leads to a regress. As long as we have an ideological part of our metaphysics, and an ontological part, then we will need to have some way of weighing up ontological components against ideological ones.

We cannot get away from the idea that there is more than one theoretical component in a metaphysical theory, with which we can measure the parsimony of the theory. And depending on which component we use to measure, we will get a different result about how parsimonious the theory is. If we knew how to weight ideological and ontological parsimony, the right rate to use when trading them off against each other, then we could come up with an objective answer about how parsimonious that theory is. But as it stands, we have no such exchange rate, and so how parsimonious a theory is will always depend on what kind of parsimony we have in mind.

Schaffer (2014, p. 2) points out there could be many levels to our ontology, with one level being the most fundamental. If so, that is the only level which we have to worry about. For example if we are measuring parsimony by looking at the number of entities posited eg forces, fields, particles, then adding a new particle will not make any difference, whereas if we are counting at a lower level including different kinds of particle then clearly it will. (This view can be contrasted with what he calls a flat ontology.)

The question is whether this general problem is an issue specifically when we are dealing with the issue of reducing causation. Recall why a reductive analysis is supposed to give us a more parsimonious theory. The theory in question refers to our overall metaphysics, not to the analysis of causation by itself. The reductive analysis is supposed to give us a more parsimonious theory because when we list our ideological primitives, we don't have to list causation. We need only list whatever it is we have reduced causation to (perhaps spatiotemporal contiguity, perhaps facts about regularities, or maybe similarity relations between possible worlds). It looks like we are

only talking about the ideological aspect of our metaphysics here, not the ontological aspect. Unreduced causation, if it is part of our metaphysics, is an addition to our ideology; it isn't as though advocating unreduced causation would mean adding special things called 'causes' and 'effects' to our list of what exists. The events which are causes and effects exist either way; the question is how they relate to each other. So while there seems no way to diffuse the exchange rate problem, perhaps it doesn't apply to the question of reducing causation. Providing a reductive analysis of causation seems to decrease the number of ideological primitives which we need in our theory, and doesn't seem to involve a trade-off, or some ontological price for this gain of ideological parsimony.

In the following section, however, I will show that the problem with measuring parsimony makes it hard to justify why parsimony should be a theoretical virtue. I will also argue that even if we don't feel that there is a problem with measuring parsimony, we might still struggle to find a justification for preferring the simpler theory.

8.3   Why prefer parsimony?

The question is whether we should prefer parsimony as a theoretical virtue; that is, whether we should go for a more parsimonious theory because that makes the theory more likely to be true. It's possible that more parsimonious theories are better because they are easier to understand or to work with, but this isn't the question we are considering here. Hall (2004) writes that the only reason to look for a parsimonious theory is the same reason as starting

to look for your lost keys in the lightest part of the room. There isn't any reason to think that that is where your keys are, but if they do turn out to be there you will find them quickly. In the same way, there is no reason to to think that parsimonious theories are more likely to be true, but if they are true we will have a quicker job setting them out than if we start out assuming that the true theory will be complicated.

However, this doesn't seem like enough to account for our preference for simple theories. Below, I will briefly discuss some other factors which might explain our preference for parsimony. Some of these explain our preference without justifying it. Other points will, if true, turn out to be good reasons to value parsimony. This will be a brief survey in which I will aim to introduce each of the possible reasons for preferring parsimony.

8.3.1 Reasons which, if true, would explain but not justify our preference for parsimony

First of all we have two explanations, both due to Quine, which would explain why we value parsimony in theories without justifying it.

We spot simple things and miss complex ones. Quine (1963) points out that this doesn't just go on at the level of perception, where we notice simple things more easily. It is also present in experiments. The structure of experiments can often mean that we can only get evidence of uniformity. That is, we might repeat an experiment to see if our results are reproducible. If they are, we might take it that this means the previous results were correct. But if we get different results that does not tell us anything (Quine 1963, p. 104).

(This seems like a rather dated way of looking at experiments; according to the falsificationist a result which differs from what we hypothesise can be very informative.)

The second point is that simple hypotheses are more likely to be confirmed, according to Quine (1963). He gives the example of an experiment where we get a numerical result, 5.2. If we then get a second result, 5.23, we will probably take that as confirmation of the original result. But if we recorded our original result as 5.22, then 5.23 would falsify the result. The simpler prediction is more easily confirmed and less easily falsified. This, Quine argues, is partly how we come to believe it. This view does depend on a concept of simplicity where an result with two decimal places is simpler than a result with one decimal place. This sounds plausible and yet I do not know of a justification for it. Similarly, this view could be challenged on the grounds that a result is not an hypothesis, and it is the simplicity of hypotheses which we are discussing here. I will not explore these issues in detail here, but it does not seem that this point accounts for our preference for simple hypotheses.

8.3.2   Reasons which, if true, would both explain and justify our preference for parsimony

Next, I will move on to points which, if successful, both explain and justify our preference for simplicity.

One obvious reason why we might think a more parsimonious theory was more likely to be true would be if we thought that the world were simple. If

the world is simple, simple theories are more likely to be true of it[1]. This has some intuitive appeal. After all, lightening strikes the highest point and water always runs downhill; it does seem as though nature never does more work than it has to. And if that is vaguely plausible, it also seems likely that nature would not do extra work in the metaphysical sense, for example by there being atoms and the things that atoms make up. Maybe we see both physical and metaphysical simplicity in the world around us.

However, the idea that the universe is simple is a mistake. Even a quick glance at our best scientific theory about almost any part or aspect of nature will show the it is complex and non-obvious. Assuming current theories are near the truth, that suggests that nature is not simple.

Another question would be: If nature is simple, what would it look like if it were more complicated? In the physical case, this seems fairly unproblematic. We can posit another logically possible world where the laws of nature are different, compare, and make a judgement about which world is more simple and which is more complex. However, there are some metaphysical cases, especially involving modality, where it is hard to find a point of comparison.

_____

[1]Probably. We could question, however, whether the property of simplicity applied to theories and the property of simplicity as applied to the world or portions of it are really the same property. Perhaps we use the same word just as a simile, as with big buildings and big ideas. If the properties are not the same, there seems little reason to think a simple theory is more likely to be true of a simple world, anymore than a big idea is more likely to be true of a big building. Even if it is the same property of simplicity in both cases, we could still question whether ensuring the theory has that property makes the theory more likely to be true. For example, suppose I am a scientist studying red giant stars. I will not guarantee that my theories are more likely to be true by printing my papers in red ink. So before saying that simple theories are more likely than complex theories to be true of a simple world, we would have to say a lot more about the property of simplicity, as well as the relationship of theories to the world (reference? correspondence? truthmaking?) I will not deal with this issue here, however, as it seems that the world is not simple, and what would be likely to be true of it if it were is therefore moot.

Suppose we are considering possible worlds realism and want to weigh up how simple or complex it is. We cannot simply compare a world where possible worlds realism is true to one where it isn't because possible worlds realism, if true, is true at every world. We cannot use the philosophical device of possible worlds to measure up the simplicity or otherwise of theories about possible worlds. We will run into the same problem if we want to measure up the simplicity of other claims with modal content. For example, I could weigh up the simplicity of the view that God exists, by comparing a world with God to another possible world without God. But I would not be able to do the same if the claims being evaluated are 'God necessarily exists' and 'God necessarily does not exist', since both of these, if true, are true at every world. So at least in the case of some metaphysical hypotheses, we will struggle to measure simplicity due to having nothing to compare against.

Another reason to prefer simplicity would be that, in our past experience, we have pursued simple theories, and these have turned out to be true, or at least useful. The argument goes that we can tell, from our past successes, that the principle of looking for the simplest explanation is a good one, even if we do not know why.

A variant of this view might be that, in the past, the principle of parsimony has worked particularly well for science, and philosophy should therefore adopt it. Huemer (2009) suggests this view in his paper 'When is Parsimony a Virtue?'.

The problem with both these views is that they are circular. The simplicity of current theories is part of the reason we have for believing them true. Thus, their truth cannot also be a reason to believe in the principle of

parsimony, on pain of circularity. How much of a worry this is will depend on to what extent parsimony was a reason for adopting our current theories in the first place. If other considerations motivate our adoption of our current theories, and all those theories just happen to be simple, then that would provide evidence that truth and simplicity at least tend to go hand in hand.

However, there is also the point that what works for science may not work in philosophy (Huemer 2009). Only in science do we have a body of theories which are widely accepted as almost certainly true. In philosophy there is no such consensus. So it is much easier to draw conclusions from our scientific theories along the lines of 'all our true theories have had this quality, so this quality must be truth conducive'. We have no body of philosophical theories that we can draw conclusions from in the same way.

Another possible justification for our preference for simplicity might be based on the number of simple and complex theories. Huemer offers two possible accounts here: the boundary asymmetry account and the numerousness account. I will discuss these together, since I feel that objections to one of them will probably turn out to be objections to the other one too. The boundary asymmetry account is as follows. Imagine first of all that we can list all possible theories and rank them in order of simplicity. Huemer claims that there will be a most simple theory but not a most complex. The list of possible theories is then bounded in just one direction; there is a lower bound on simplicity but no upper bound. The probability of each theory will then decrease as the theories get more complex, with the most simple theory being the most probable. Huemer argues that to assume the most probably theory was any other theory besides the most simple would be arbitrary.

The boundary asymmetry account is based on the assumption that there will always be a common measure of simplicity. I have argued above that this is not true. There may not always be a common measure of simplicity: a unit where having more of it makes a theory more complex. Some theories involve entities. Some involve parameters. Some involve the way things relate to each other. The simplicity of theories cannot always be quantified or compared in a way which is consistent or meaningful.

Even if someone were to argue that there is some consistent and objective way of ranking theories by their simplicity, we could still challenge these accounts on other grounds. Below, I will present counterexample to the idea that, if we can rank hypotheses from the most simple to the most complex, there will be a most simple but not a most complex. Throughout this example we can keep an open mind about how simplicity ought to be measured; or, if you prefer, you can pick a specific measure of simplicity. Whichever measure you pick, there will be a most complex hypothesis.

Suppose Alice rolls 120 on a matching set of one or more fair numerically balanced dice. Bob is coming up with hypotheses about how Alice rolled that number. He does not know how many dice she rolled or how many sides these dice have. Obviously one possibility is that she has 60 d2 dice (coins), each of which showed a 2. Another possibility is that she has just one d120 dice, which showed 120. Another possibility would be that she has six d20 dice, each of which showed a twenty. And so on.

First, let us run the argument with the measure of complexity being the number of dice (vaguely plausible; the dice are the entities involved after all). Then the simplest hypothesis will be that Alice has one d120 dice. But then

there will also be a most complex hypothesis: that Alice has 120 d2 dice.

Next, let us run the argument with the measure of complexity being number of sides of the dice (vaguely plausible; they invented coins before d120 dice). Then the simplest hypothesis will be that Alice has 120 d2 dice. In that case the most complex hypothesis is that Alice has a d120 die. The reason for this is that no larger numerically balance dice are possible. This fact is grounded in the geometric and mathematical nature of dice. Within the field of hypothesising about what dice someone has, then, there most certainly is a most complex theory as well as a most simple.

This is a counterexample to the idea that there is always a most simple hypothesis but not a most complex. As an aside, this is also a fairly good illustration of my previous point that simplicity is hard to measure. There does not seem to be anything to choose between measuring complexity in number of dice and measuring complexity in number of sides. However, if someone did feel that there was a definite answer about which, dice or sides, was the correct measure of complexity, my main point would still stand: either way, there is a most complex as well as a most simple.

Now there are two obvious objections here. One is that I am talking about an example of a theory in an artificial example field. Huemer is talking about theories in very general terms, separate from the subject matter. I would argue, however, that theories are theories about something. And that subject matter limits the possible theories available to us. There are not an infinite number of theories about how many dice Alice has and how many sides they have, and the subject matter gives us a most complex. I suggest that this will turn out to be the case in some real life cases of philosophical,

mathematical and scientific theories. Certainly, there are plenty of cases where there is a most complex as well as a most simple. Examples include: the platonic solids, allotropes of carbon, stable elements. Even if we can rank hypotheses in order of simplicity, there may well be a most complex hypothesis as well as a most simple.

The numerousness account (Huemer 2009), runs as follows:

- P1: All theories about a given thing can be divided into two classes, the simple and the complex.

- P2: Each class is equally likely to contain the true theory.

- P3: Complex theories are more numerous

- C: Therefore, simple theories are on average more complex

I have already suggested that it is not possible to rank hypotheses in order of simplicity. If this is the case then dividing theories into two classes, simple and complex, is going to be challenging. However, I want an argument here that will work even for readers who believe that hypotheses can be ranked in order of simplicity.

P2 can also be challenged on the grounds that there is more than one theoretical virtue. Even if parsimony is a theoretical virtue, there are clearly others: most obviously, explanatory adequacy. As well as being simple, a good theory will account for the evidence. I think it is plausible that explanatory adequacy is more important than parsimony (even if parsimony is a theoretical virtue at all). There is no reason to suppose that parsimony is the first criterion we would apply when choosing an hypothesis. In fact

it seems obvious that we don't: when we generate an hypothesis we are looking first of all for something that would explain the evidence. Parsimony might then be used to choose between hypotheses which explain the evidence particularly well. With this in mind, it doesn't seem right to say that probabilities are equally spread over all the possible hypotheses. Some hypotheses are more likely to be true because they explain the evidence more effectively. There is no reason to believe that these are just as likely to be in the more complex category as the simpler category. The numerousness account presupposes that parsimony is the first theoretical virtue which we would apply when selecting hypotheses; we have no reason to think this is true, and some reason to think that it isn't.

With these thoughts in mind neither the numerousness account nor the boundary asymmetry account are effective ways of arguing for parsimony as a theoretical virtue. They do not justify a preference for simplicity. Both depend firstly on the belief that hypotheses can be ranked in order of complexity and secondly on taking the hypotheses out of context and ignoring what it is that they are hypotheses about. Either of these reasons is enough to reject the numerousness account and boundary asymmetry account.

The last possible justification for our preference for simplicity which I consider here is another suggestion from Huemer. He refers to it as the likelihood account. He argues that simpler hypotheses are less likely to be confirmed. He offers an in depth explanation of how this is so, but perhaps we do not need to examine this in detail. Instead let us notice that fact that Quine (1963) says the exact opposite: simple theories are more likely to be confirmed. The reason why they make opposite claims is this: they are using

different measures of simplicity. Quine is talking as though the hypothesis is a number (a result) which is more complex if it is to more decimal places. Huemer is talking as though simplicity is measured in parameters, where a hypothesis is more complex if it has more parameters. It may well be true that, on Quine's conception of simplicity, simple hypothesis are more likely to be confirmed and, on Huemer's conception of simplicity hypotheses are less likely to be confirmed. The fact that two philosophers can talk about simplicity and get opposite results because they measure it in different ways, drives home the point that we lack a universal measure of simplicity and talking about simplicity and complexity outside of the contexts of the hypotheses involved is not fruitful.

However, someone might respond by saying that there is a good measure of simplicity, and that Huemer is correct about what it is: a more simple hypothesis is one which has fewer parameters. In that case Huemer's likelihood account may justify our preference for simplicity in science. But as Huemer points out, this does not mean that we have a reason to think that parsimony is a virtue in philosophy. Philosophical hypotheses generally do not have parameters. Philosophical theories tend to be questions about how certain things relate: what is identical to what? What grounds what? rather than questions about what the parameters governing some process are. So the likelihood account may, depending on our other commitments with regard to simplicity, justify a preference for simplicity in science. But it does nothing to justify it in philosophy.

8.4 Conclusion

In conclusion, let us remind ourselves of the context here. We are talking about whether parsimony is a reason to provide a reductive analysis of causation. In chapter one, I have argued that reductive analysis is best understood as the discovery of a grounding relation. In particular, I have argued that when we discover that *a* can be reductively analysed in terms of *b*, that does not mean that *a* is identical with *b*. So if we can provide a reductive analysis of causation, that does not mean that we are positing fewer relations in total. The relation of causation will still be present in our ideology, as well as whatever relations we have analysed it in terms of. So we will not have achieved simplicity in the form of a smaller ideology at all. However, if we discover a grounding relation between causation and our analysans, then we will have shown that causation is not fundamental. So, we will have a fundamental ideology which is smaller overall. Schaffer (2014) writes that only fundamental entities (and presumably properties and relations) need count from the point of view of parsimony. Derivative entities are a free lunch. So from this point of view, producing a reductive analysis of causation could lead to fewer fundamental ideological components in our metaphysics.

But we have seen above that the only justification of parsimony as a theoretical virtue is based on the notion that we are measuring the simplicity of an hypothesis in terms of number of parameters. It makes sense in science, but not in philosophy. This justification doesn't translate into a reason why we should look for a metaphysics with fewer fundamental ideological components. We have been given no reason at all for preferring more parsimonious

theories in the context of metaphysics, and so parsimony is not a good reason to look for a reductive account of causation.

Chapter 9

Empiricism

9.1 Introduction

In this section I am going to suggest that empiricism is a good reason for wanting a reductive analysis of causation. Roughly, the argument runs as follows:

> P1: EMPIRICISM: We cannot know about anything except through our senses.
>
> P2: KNOWLEDGE OF CAUSATION: We know some causal facts.
>
> P3: HUME'S THESIS: We cannot know about causation through our senses.

We can already see that this leads to a contradiction. Empiricism and Hume's thesis between them contradict P2. We end up saying that we both can and cannot have knowledge of causation. This is the problem. We are very strongly inclined to believe that we do have at least some knowledge of

what causes what. But empiricism seems to contradict that. In the first part
of this chapter I will explain P1 and P3 in more detail.

We can alter the argument to avoid the contradiction, by replacing P1
with:

> P4: EMPIRICISM': We cannot know about anything except through
> our senses, unless it can be reduced to something we can know
> about through our senses.

We need a reductive analysis of causation, specifically one which reduces it
to something we can know about through our senses. The point of reduction
is not parsimony, but explaining how it is possible that we know any causal
facts. The reason for reduction is epistemological rather than metaphysical.
Later in the chapter I will talk more about P4, and how reduction is supposed
to help us know about something.

First, however, I want to tackle an obvious objection up front.

9.2   Causation is not a theoretical entity

Here's one very obvious worry about what I said above. If we look at the
premises, we could not substitute anything else for causation. We do not
argue in this way about anything else.

Take something else that we think we can know about, for example atoms
or electromagnetism. We do not directly perceive these things through our
senses. So far, this is like causation. But we do not look to reduce them
to things we can perceive through our senses. In fact, the reduction goes

the other way. The perceivable things are macroscopic objects. We think atoms are more fundamental, so if there's a reduction to be done, it's in the direction of reducing macroscopic objects to atoms. And we think the properties of atoms explain the properties of the macroscopic object (the salt crystals are cube shaped because of the molecular structure, which in turn depends on what is going on with the atoms). So what I am suggesting in the case of causation, that it be reduced to something more fundamental and more empirically accessible, is different from our approach to other theoretical entities.

The answer to this is partly that causation is not like atoms or electromagnetism. It is empirically inaccessible in a different way. Certainly we cannot see atoms, or electromagnetic forces. However, they are objects in a theory. There have, at least in the past, been competing theories, and evidence was brought to bear suggesting that the theory involving atoms (as we now understand them) was the right one.

Causation is not like this. It is not an object in a theory [1]. It is not clear what a competing theory even could be. Possibly the competing theory would simply be that things happen one after another with no causal links between them. But it is not clear what evidence could be brought to bear in favour of the idea that causation links events. This is perhaps a more complicated way of putting Hume's point: we do not see causation, and we also do not see evidence for or against causation. To have evidence for or against it, we would have to be able to say what the world would look like if

---

[1] Debatably. The Canberra Plan is one possible approach to causation, and does involve treating causation as an object in a theory; Menzies (1996) advocates this approach. However, in chapter six I argue that causation is not a theoretical entity.

it wasn't causal, and then say that it doesn't look like that. Hume's point is that the world would not look any different. Perhaps in the light of this we should replace P3 with

> P5: MODIFIED HUME-LIKE THESIS: Causation is not part of a theory which we can confirm or falsify with evidence from our senses.

So when we say that we cannot know about causation through our senses, we really mean something further than that: we don't really know what would count as evidence for or against the thesis that the world as a whole is causal, that one event can cause another event.

Furthermore, examples of accounts which aim to reduce causation to something empirically accessible do not always reduce it to something we can see. The most obvious example is Dowe's conserved quantity account of causation (Dowe 2000). It certainly isn't the case that a conserved quantity of energy or momentum is empirically accessible in the straightforward sense of being able to see it. However, conserved quantities of energy are theoretical entities in a theory which empirical evidence can confirm or falsify. So reducing causation to these theoretical entities helps to make it more empirically accessible. We still don't see it, but we have grounded it in something for which there can be evidence.

9.3   Empiricism

Above, I have defined empiricism as the view that we cannot know about anything except through our senses. However, in the previous section I have

suggested that the range of things which are empirically acceptable is not co-extensive with the range of things we can see. Another way for something to be empirically acceptable is to be part of a theory where some evidence can be brought to bear for or against it.

To illustrate this, we can consider a comment attributed to Wittgenstein. He and a friend were watching the sun set, and his friend commented that he could see why the sun was once thought to go around the earth. Wittgenstein asked why anyone would have thought that, when in fact the earth is rotating. His friend replied that it just looks as though the sun goes around the earth. Wittgenstein replied, 'Really? Then what would it have looked like if it had looked as though the earth was rotating?' In order for some evidence to count for or against a theory, we have to know what things would look like if that theory were not true, or to put it another way, what evidence would falsify that theory.

According to this approach, no one claim is confirmed or falsified by evidence. Rather, the whole theory made up of all the claims we consider to be true, will be confirmed or falsified by evidence. Deciding when there is enough evidence that we should make changes to our entire theory is a complex business which will involve considerations of conservatism and simplicity. This sort of view is known as epistemological holism, and is due to Quine (1951) and Duhem (1954). This view of empiricism is widely accepted and a plausible approach; it is the view I will be using here.

9.4   Hume's thesis

The most familiar empiricist worry about causation is due to Hume (2000 [1739]). Hume's empiricism was a part of his philosophy of thought and perception. He argued that our minds contain ideas and impressions. We get impressions directly from the world around us, while ideas come from impressions, either on their own or combined. Ultimately, every idea that we have must come from a corresponding impression or impressions, according to Hume. So, for each idea, we can ask where we got it from: what impressions gave rise to it.

Part of the notion of causation is the idea of necessity. The cause doesn't just come about before the effect, but rather necessitates it: it makes it happen. But, Hume argues, it is not clear how we could have ended up with an idea of necessity. We never have an impression of necessity. What we see is one billiard ball striking another and the second one moving off; we don't see the necessity of the second ball's moving when struck. Nor can we get the idea of necessity by combining other impressions. The best explanation, Hume argues, of our apparent idea of necessity is that when we see E follow C very often we form a habit of expecting the one to follow the other. But this doesn't mean that we know that E will always follow C, or that we really have an idea of necessity, just a mental habit. Hume's thesis, then is that we can have no idea of causation because we have no impression of necessity.

In my argument I use a modified version of Hume's thesis according to which something is empirically accessible if it is part of a theory where evidence can bear for or against it.

9.5 An objection: seeing causation

One way of objecting to my argument would be to deny Hume's thesis and my variant on it. Some writers argue that we can simply see causation, and so none of this is needed.

Anscombe (1993) argues that we directly see particular cases of causation. After all, we use a variety of causal verbs:

> *scrape, push, wet, carry, eat, burn, knock over, keep off, squash, make* (eg noises, paper boats) *hurt...*
> (Anscombe 1993, p. 93)

and we learn how to apply them from observing the world around us. She argues that those who argue that we cannot see causation have set the bar unrealistically high for what would count as really perceiving causation, or that this has been artificially excluded from the list of kinds of thing we can perceive. She considers it obvious that we do perceive individual cases of causation, and someone who disagrees would never consider anything to be a case of seeing causation (Anscombe 1993).

The argument in its simplest form is that most people would say they have seen causation when they observe one billiard ball hitting another. There is more to Anscombe's argument than this, however. She argues that the necessary connection is not part of causation; causation is indeterministic. This is not supported by an appeal to quantum mechanics but, much more prosaically, an example of catching an infectious disease. If I'm exposed to someone with an infectious disease and I then get the disease, I will say that the exposure caused the disease. But, Anscombe points out, just because

I have been exposed does not mean I will catch the disease. Effects come from their causes, but there is no necessary connection. Anscombe takes a singularist view of causation; the particular case of causation is fundamental. We do not have to worry about necessity, which as a modal concept seems unobservable, or regularities, which would be observable only if we had time to observe the entire universe. We just have to look at the instances of causation which are under our noses. In a similar vein, Strawson (1985) argues that causation is something which we can directly experience, when we are acting or being acted upon.

However, there might still be a need for a reductive analysis even if we do see causation. The problem is that seeing causation might not be enough for it to be empirically acceptable (in fact seeing causation would be neither necessary not sufficient).

It is not always clear what the difference between observing and inferring is. People can look at the same event and see very different things. People bring their assumptions with them they perceive things, so that their perceptions differ from each other. For example, if you see a black and white picture of Santa Claus and are asked later what colour his coat was, you might well say red. You might even forget that the picture had been black and white, and recall red because you know that Father Christmas wears a red coat. Similarly, people with colour-grapheme synthaesthesia see letters as having colours, but not the same colours as each other: one person might see 'A' as red, while another might see it as yellow. In other words, our prior knowledge, assumptions and neurological conditions in our brains make us see different things.

So even if we do see causation, this might not mean anything significant in terms of causation's empirical status. We see it because we expect to see it. We see colour as well, but we can still tell a story about light, wavelengths, reflection and absorption in order to explain how we see colour. On the one hand we might perceive a white object as red because it was under a red light. On the other hand, we have the case where we perceive a red object as red. We see colour: sometimes we see the 'real' colour, sometimes we see the wrong colour and sometimes we see colour where there is none. This isn't why we regard the notion of colour as empirically acceptable. It is because when we examine the nature of colour more closely, we find that there is an underlying story about wavelengths of light which explains why we see certain things as certain colours under certain conditions, and we can imagine evidence which would falsify this theory.

We know we can see colour, but we need more than that for it to count as an empirically acceptable concept. We need an explanation of it in terms of wavelengths of light, which are not merely empirically observable, but are scientific concepts subject to scientific scrutiny.

Even if we know that we see something we can still ask for a story about how we come to be seeing something. Even if we do see causation, we can ask how it is possible that we see it. I definitely can see ordinary objects like tables, and yet we could still ask how, and give a story about how I see tables, including something about the atoms making up the table and the light which reflects off it. What we want is the corresponding story about causation. For empiricism, it is not enough to just see causation; we need to know how we see causation and have some reason to think that we are seeing

something real.

9.6   How does reducing causation make it empirically acceptable?

I have argued that there is a need for a reductive analysis of causation, since directly seeing causation is not enough to make it empirically acceptable. I now need to show that a reductive analysis of causation at least has the potential to secure what we want. We need it to be the case that if you take some concept x and reductively analyse it using other concepts y, z and w which are empirically acceptable, then you have shown that x is empirically acceptable too.

This is now an epistemological question about knowing things empirically and about the nature of reduction. I discuss the relation of reductive analysis in detail above in chapter 1. What we need is for the following statement to be true:

P5: If I empirically know things about y, and I have a reductive analysis of x in terms of y then I can empirically know things about x.

9.6.1   Knowing that I have seen Phosphorus

'Know' is an intensional transitive verb. That is, the truth value of claims containing that verb can change if the object of the verb is swapped out for a co-referring term. For example, suppose I can say truthfully 'I know I have

seen Hesperus'. It still might not be true that I know I have seen Phosphorus, even though Hesperus and Phosphorus are one and the same, if I don't know that 'Hesperus' and 'Phosphorus' refer to the same object.

This principle can also be useful when we consider causation. Suppose we adopt an empirical analysis such as due to Fair (1979). Roughly speaking, we are saying that causation can be reduced to the transference of energy or momentum. Suppose we know that momentum was conserved in the interaction between two billiard balls, for example. But we don't know that causation has taken place. If we then discover that causation is the conservation of momentum or energy, then we know something more than we did before.

If to reductively analyse a concept were to discover an identity relation, this would be a very simple story. We have empirical access to the objects we analyse causation in terms of, and discover an identity between that and causation, and then of course it follows that we have empirical knowledge of causation. However, the situation is more complicated than this, because as I discuss above in chapter one, analysing something is not the same as discovering an identity. I think that analysis is still close enough to identity for us to get the right results however.

9.6.2 Analysis is not identity

In chapter one above, I argue that to reductively analyse a concept is to discover what grounds it. Grounding is the non-causal dependence of something less fundamental on something more fundamental. A good example of

grounding is that Socrates grounds the set singleton *Socrates*. The relation is close to identity, but more directed and explanatory. The existence of Socrates explains the existence of the set singleton *Socrates*.

An analysis of causation such as Dowe's is much the same. It is almost, but not quite, an identity. Causation is grounded in and depends on the more fundamental conservation of momentum or energy. Wherever there is causation there is a conservation of some conserved quantity of energy or momentum, and vice versa[2].

So, although reductively analysis is not identity, the analysans and the analysandum will have the same extension. This means that once we have made the analysis, and discovered a case of the objects in the analysans, we can infer that the analysandum took place too. In this respect, making a reductive analysis of causation is just like discovering that Hesperus is Phosphorus.

9.6.3   Knowing empirically and knowing

So far my argument explains how reductive analysis could allow us to know something about causation when we did not before. What we particularly want though is for our reductive analysis of causation to ensure empirical knowledge of causation.

Most of the burden here is on the actual analysis of causation, to be entirely in terms of something which is empirically acceptable. An analysis such as one of those due to Dowe (2000), Fair (1979) or Salmon (1984) will meet this criterion. Causation is analysed entirely in terms of things which

---

[2]Not forgetting that Dowe believes inertia is causal.

are part of some theory where evidence can bear for or against it.

Salmon (1984) offers a mark transmission theory of causation. Very briefly, a causal process is a process which transmits a mark. A mark is an alteration to a characteristic produced by a single interaction. Evidence can bear for or against this in the sense that we can inspect whether an object has or transmits a mark. It is a non-modal, observable criterion. Fair (1979) offers an account of causation in terms of transfer of energy or momentum. Again, this is actual and observable, and is essentially a reduction of causation to processes which are physically unmysterious and understood in science. Dowe's theory is similar to Fair's, but remains neutral on more details: in Dowe's theory the energy or momentum need not be transferred, but some quantity needs to be conserved. Dowe (2000, p. 110) regards this as an advantage as it avoids some of the objections to Fair's account. Dowe's theory relies on the same scientific theories as Fair's, which again, evidence can count for and against.

If analysis is the discovery of a grounding relation is an empirical discovery too, then it would seem pretty clear that any knowledge resulting from the analysis is empirical knowledge. It is not really clear that discovering a reductive analysis of causation is an empirical discovery. It is not clear that there is evidence which bears for and against it. Rather, it seems as though we have found something empirically acceptable which has the same extension as our concept of causation, and could plausibly be what grounds it. However, we can still consider our knowledge of causation to be empirical even if the discovery of the analysis is not itself empirical. Many parts of our generally accepted body of empirical knowledge rely on deductive logic

as well as empirical facts. This is no barrier to those theories being empirical overall.

A non-reductive theory of causation such as Woodward's, however, does not reduce causation to anything empirically accessible or non-causal. If we can reduce causation to some subject matter about which we have a theory, and that theory is one where evidence can bear for or against it, we have explained how we can have empirical knowledge of causation. However, if a theory does not reduce causation to something non-causal we are left with no explanation for how it is possible for us to know anything about causation empirically.

9.7   Conclusion

Empiricism gives us a reason to want a reductive analysis of causation. Hume argued that we never see causation. More than that, causation is not an object in a theory for which evidence can bear for or against. Arguments that Hume is wrong, and that we do in fact see causation, do not help us here, as seeing by itself does not make a concept empirically acceptable. But, by providing a reductive analysis of causation we can show how it is grounded in something empirically acceptable. This explains how it is possible for us to have us to have empirical knowledge of causation.

Part IV

Two kinds of causation

Chapter 10

Pluralist accounts of causation

10.1 Introduction

In this chapter I will explore causal pluralism; the view that there is more than one causal concept. Pluralism is a natural option to look at for causation, since nearly every account of causation involves more than one causal concept. The orthodox view is to settle on one concept which is causation and count the other concept as related but not quite deserving to be called 'causation'. Causal pluralism is the view that that both of these concepts are concepts of causation.

I will first look at pluralism generally, beginning with a brief description of how the term 'causal pluralism' is used, suggesting that causal pluralism is the view that, fundamentally, there is more than one kind of causation. I will move on to what motivations we might have for seeking a pluralist account of causation. I will then raise an important question about pluralism: if we have two concepts of causation, then we can ask what those two concepts

have in common which make them both causation. On the one hand, if the two concepts have nothing in common, we have no reason to use the same word. On the other, if they have something in common, we would expect that thing which they have in common to be the basis of our analysis.

I will then move on to discuss Hall's pluralist account of causation. This will allow us to see how a pluralist account will pan out in more detail. I will begin by explaining Hall's motivations for a pluralist account: he argues that there are five theses which are all true of causation, but which are inconsistent. The solution, he argues, is that the five theses are all true of causation, but not of the same causation; there are two concepts, with two of the theses belong to one and three to the other. I will consider some objections to Hall's argument and argue that, while he is right to suggest that there are two concepts, the two concepts he suggests are the wrong two.

Hall's account is of particular interest since the two concepts of causation which he posits track, or appear to track, two different kinds of causation in the literature. Some philosophers produce accounts of causation which are accounts of Dependence, and some produce accounts of Production. Illari et al. (2014, p. 255) suggest that we can use this distinction to categorise extant accounts of causation, with probabilistic causality, counterfactual accounts and manipulation and invariance accounts grouped together under difference making, and process accounts, mechanism accounts and information accounts grouped together under production. Because of this, I will describe Hall's account in detail. I will argue that the distinction Hall describes is an important one, although I disagree with his characterisation of the two concepts.

10.2 What is causal pluralism?

Pluralism about causation is the view that there is more than one kind of causation. In some ways, the idea that there is a plurality of causal concepts is completely uncontroversial. For example, Dowe defines causation by way of first defining causal processes and causal interactions. This does not amount to pluralism, because in this case the writer thinks there is just one concept which is really causation, but the best way to define it is just to start with other related causal concepts.

However, most writers define general causation ('smoking causes cancer') in terms of singular causation. Lewis defined direct causation first, and then took the ancestral to make it transitive, giving indirect causation (Lewis 1973b), (Lewis 2004a): again employing two causal concepts. In this case both general and singular causation count as causation, and direct and indirect causation also count. However, we do not think of either of these as pluralism because in these cases one of the concepts is more fundamental than the other. For example, for the singularist, singular causation between particulars is fundamental, and general causation between kinds of event is derivative.

In other words, a theory only counts as pluralist if it posits more than one kind of causation, and neither account is more fundamental than the other. Another way of putting it would be this: if we offer a univocal analysis, we are saying that causal facts are grounded by some other kind of non-causal fact. But if we are causal pluralists we are saying that some causal facts are grounded in one kind of fact, and some in another.

Causal pluralism is different from having a disjunctive theory of causation. In the case of causal pluralism, we have two concepts, or two kinds of causation, while in the case of disjunctivism, we have one kind which is disjunctive. If we offer a disjunctive analysis of causation, we are saying that causal facts are grounded either in one kind of fact or in another, but not committing to which it is. This is different from causal pluralism, where we are committing to saying that there are two (or more) kinds of fact which ground causal facts, but where for any given causal fact we could in principle say which kind of fact grounds it.

This is a tricky distinction. We might think that disjunctivism and pluralism amount to the same thing for this reason: surely we could turn any pluralist account into a disjunctive account by simply putting the analyses of the two concepts of causation together, and connecting them by a disjunction. But to do this would be to make a different claim. In the case of pluralism, there are two things which deserve to be called causation. In the case of disjunctivism, there is one thing, but that thing is described disjunctively. In a disjunctive analysis, neither of the disjuncts is causation. Rather, to be causation is to be either the first disjunct or the second disjunct.

Since a disjunctive analysis is by definition a failed analysis, there are not many examples in the literature. On example is the toy analysis Lewis (2004b) offers in 'Void and Object' as an example of why he believes analysing causation in terms of biff will not work. He suggests we would have to say that occurs when one of a number of distinct conditions were satisfied. Lewis's toy example goes like this:

1. Event c directly causes event e iff c stands to e in the relation that occupies the biff-role. For short: iff c biffs e.

2. The absence of any event of kind C directly causes event e iff, had there been an event c of kind C, c would or might have biffed some event d incompatible with event e.

3. Event c directly causes the absence of any event of kind E iff c biffs some event d incompatible with any event of kind e. of

4. The absence of any event of kind C directly causes the absence of any event of kind E iff, had there been an event c of kind C, c would or might have have biffed some event e of kind E.

(Lewis 2004b, pp. 284-5)

It is not that each of these disjuncts represents a kind of causation. Rather, each one is required in order to count as causal all the cases which should be counted as causal.

If we consider a disjunctive analysis to be not a kind of a analysis but a failed analysis, we could still be a pluralist; we would just need to say that there is no property of being either one kind of causation or the other. Indeed we might take this approach if we felt that our analysis was bound to yield something unnatural. Rather than end up with a disjunctive analysis, we might say that there were really two kinds of causation all along, and that our initial attempts to give a single analysis of one concept of causation were misguided.

10.3   A pluralism puzzle

The idea of a pluralist analysis leaves us with a dilemma. Suppose we have pluralistically analysed causation into causation$_1$ and causation$_2$.

On the one hand, it might be the case that causation$_1$ and causation$_2$ have nothing in common, except that we use the same word for them. In this case, causation is like 'mean' (average) and 'mean' (harsh): it simply has two separate meanings which have nothing to do with each other. This is possible. But it seems unlikely given that if 'causation' does have two separate meanings, we use them in such similar contexts. After all, 'mean' and 'mean' are used in the quite different contexts of maths and human behaviour, so we have no trouble knowing which sense is meant. There would be no such clarity with causation$_1$ and causation$_2$. In a similar vein, everyone who uses 'mean' and 'mean' knows about the two meanings (three meanings, in fact), and knows which one they are using at the time, but this wouldn't be the case with causation$_1$ and causation$_2$. Finally, if it is the case that we just have a homonym here, it would make sense to analyse causation$_1$ and causation$_2$ entirely separately, as two different projects. But while all this seems possible, it does not seem very likely, or indeed what writers who advocate causal pluralism are intending on endorsing.

On the other hand, it might be that causation$_1$ and causation$_2$ do have something in common, which explains why they both deserve to be called 'causation'. It would also explain why we analyse them together, as one project. But then we have to explain why we do not use whatever the two concepts have in common as the core of our analysis. There will of course be

many difference instances of causation, many of which can be grouped into kinds in various different ways. But this doesn't prompt us to offer a wildly pluralistic analysis, treating each case separately. Of course, we just look for what they have in common, for what makes them all cases of causation, and try to analyse that. So we would have to be able to say why these two concepts of causation ought to be analysed separately, rather than together using whatever they have in common.

One obvious reason for pluralism might be that no one analysis can give us the 'right' answer for all the different possible cases we can come up with. We have cases where we are fairly clear in our minds that causation has occurred, and we want an analysis which counts those as causation. On the other hand we have cases which do not sound causal to us, and we want our analysis to tell us that these are not cases of causation. So, when I talk about whether an analysis can 'cope' with a certain case or whether it gives the 'right result', this is what I mean.

So, we might adopt two analyses, so that by picking the right analysis, we can always get the 'right' answer for any given case. But this strategy seems very unappealing. The fact remains that our gut feeling about what the 'right' answer is in any given case, is only that, a gut feeling. It is hardly conclusive evidence for anything. And worse: there would be no principled way of knowing when we had picked the 'right' analysis to use for this particular case. Say we have some case and we want to know if it counts as a case of causation. Enter our analysis. This should provide us with a checklist to discover if what we have here is a case of causation or not. But if we have two analyses, with different extensions, and no overall

principle about which to use in which case, then in some cases we will still be undecided, or our decision about whether to count the case as causal or not will be arbitrary.

This use of cases to discover a correct analysis is a two way process. Whether we are prepared to accept or reject an account of causation partly depends on whether it copes with all the cases we can think of. But our response to those cases will also be influenced by our views on which account of causation is right. For example, suppose a knock at the door makes me burn the soup. The knock at the door prevents me from stirring the soup, and if I had stirred the soup, it wouldn't have burned. This might sound like a counterexample to any kind of biff analysis, or physical process analysis, because there is no physical link between the knock at the door and the burned soup. But if we are very committed to a biff account of causation, we might well be prepared to deny that the knock at the door did cause me to burn the soup. Sure, what went on is a bit like causation, and in some everyday cases we might want to treat it like causation, but it isn't really causation, precisely because it lacks that physical link.

There are certainly many cases like this where we are pulled two ways about an example. Presented in one way, we want to see it as causal, but thought of another way, we don't. A pluralist analysis of causation then, might do more than just provide us with a way to get the 'right' answer for all the cases of causation. Rather, it could explain why in some cases we are genuinely pulled both ways. When we think of causation$_1$, we want to say that the knock at the door made me burn the soup, but when we think of causation$_2$, we want to say that it didn't. In this way, a pluralist analysis

could work just like a univocal analysis, in that it can neatly explain our intuitions about a variety of particular cases, in a general principled way.

10.4 Hall's reasons for pluralism

In his paper, 'Two Concepts of Causation' Hall (2004) argues for causal pluralism. I will explain Hall's approach to causation here, since although I do not agree with all the details of his account, his argument for causal pluralism has some very compelling elements.

Hall picks out five theses which all sound convincingly true of causation, and shows how in certain cases they conflict. He argues that there are actually two concepts of causation, with some of these theses belonging to one and some to the other.

- *Transitivity:* If event c is a cause of d, and d is a cause of e, then c is a cause of e.

- *Locality:* Causes are connected to their effects via spatiotemporally continuous sequences of causal intermediates.

- *Intrinsicness:* The causal structure of a process is determined by its intrinsic, non-causal character (together with the laws).

- *Dependence:* Counterfactual dependence between wholly distinct events is sufficient for causation.

- *Omissions:* Omissions – failure of events to occur – can be both cause and be caused.

(Hall 2004, pp. 225-6)

In fact, dependence is normally taken to be sufficient but not necessary for causation. Hall's argument is that there is a tension between *transitivity*, *locality* and *intrinsicness* on the one hand, and *dependence* and *omissions* on the other. Cases of double prevention and omission bring out this tension. Hall argues that counterfactual analyses need the first three theses to deal with some examples, while in other cases, those three theses make trouble for counterfactual analyses.

Essentially, Hall's reasons for causal pluralism is that he thinks no univocal analysis of causation can do the job. In the next sections I will go on to explain why Hall thinks these five theses will come into conflict and why two concepts of causation are needed.

Hall describes the target of his analysis: a concept of causation which is transitive, egalitarian and between localized events. By egalitarian, Hall means that all causes count as causes, not just the salient ones. That is, there is no distinction between causes and background conditions. This means we will be happy to count the formation of the earth as a cause of today's bad weather, as well as similar technically correct but absurd sounding claims. Hall also explains that he will be setting out to discuss a simple counterfactual analysis, not one of the many more refined versions, to keep his argument simple. Hall also stipulates that he is discussing causation in a deterministic world, with no backwards or simultaneous causation or action at a temporal distance. The idea is that if Hall's pluralistic analysis can be made to work for this limited range of cases, it may be possible to expand it so that it

works across the board for example in cases of indeterministic causation.

I want to go to look briefly at the counterexamples Hall uses to show that the five theses come into conflict. This is crucial to Hall's argument, since it not only shows why he thinks there is a problem with all five theses being true of one concept, but why he splits it in the place that he does, with intrinsicness, locality and transitivity on one side and omissions and dependence on the other.

10.4.1 Early pre-emption

Early pre-emption cases are cases in which the effect does not counterfactually depend on the cause, because some back-up cause would ensure that the effect happens anyway. A classic example of this is two assassins. Both independently get into position to shoot the target, but assassin A shoots first. Seeing that assassin A has done his work for him, assassin B does not bother to shoot. However, if assassin A had not shot the target, assassin B would have done the job. So the target's death does not counterfactually depend on A's shot.

The way of dealing with this is to invoke transitivity. Instead of seeing the two events as 'A fires' and 'target dies', we can think of all the intermediate events which form the causal chain between the two. For example, 'target dies' depends on 'A's bullet flies towards target' which depends on 'A fires'. There is no chain of causal events between B and the target, because B does not in fact fire. So, counterfactual theories can deal with cases of early pre-emption.

Hall points out, however, that they can only do this by including the thesis of transitivity in the account. This is an important point for Hall; not only are all the theses plausibly true of causation, but they are all needed in order to cope with certain cases.

10.4.2   Late pre-emption

We have already discussed this type of overdetermination in chapter four. Briefly, these are case where, as with early pre-emption, the effect does not counterfactually depend on the cause because some back-up cause is present to ensure that the effect would have happened anyway. But unlike early pre-emption, late pre-emption does not involve a continuous causal chain from the cause to the effect which is missing between the back-up cause and effect. In late pre-emption cases, assassin B fires a split second after assassin A, so that the target's death does not counterfactually depend on A's bullet; B's bullet is also present to be a back-up cause, right up until the point when the effect occurs. This means that using transitivity and pointing to the causal chain between the cause and effect will not help us distinguish the cause and back-up cause, since the chain exists in both cases.

Hall considers that there are two promising ways of coping with these cases. One depends on Locality: we add to our analysis of causation that causes must be spatiotemporally connected to their effects using spatiotemporally continuous sequence of causally intermediate events. This analysis gives the right result for this case, that Suzy's throw and not Billy's causes the bottle to break. Using this approach would mean adding the thesis of

locality to our theory, however.

The other way to handle cases of late pre-emption is to make use of the thesis of intrinsicness. Hall takes this strategy from Lewis. The basic strategy is to say that although Suzy's throw doesn't have the right pattern of causal dependence, it is intrinsically like a process that does. The only difference between Suzy's throw in this example, and Suzy's throw in a case where the bottle's breaking does counterfactually depend on her throwing the rock, is something extrinsic to the process: Billy's throw.

Hall is unconvinced by Lewis's talk of processes, but thinks Intrinsicness can be used to rescue the counterfactual account, in the following way. Instead of talking about processes, Hall talks about a structure of events: the structure of events including effect e which happens at time t' and all its causes back to an arbitrary earlier time t. (Hall 2004, p. 236) This structure of events has an intrinsic character, which is just to do with the events and the relations between them, not about anything external to them. The Intrinsicness thesis states, then, that if two structures of events have the same intrinsic character, they also have the same causal character.

Either way, Hall argues that counterfactual theories need at least transitivity and also either locality or intrinsicness, in order to deal with pre-emption cases.

10.4.3 Double prevention and omissions

The theses of Locality, Intrinsicness and Transitivity come into conflict with Dependence when we look at examples of double prevention and omissions

as causes or effects.

Cases of double prevention are those where an event c happens, preventing d, where if d had happened, it would in turn have prevented e. For example, a knock at the door prevents me from stirring the soup, and if I had stirred the soup, that would have prevented it from burning. Hall argues that in cases like these we have conflicting intuitions about whether the knock at the door (for example) should count as the cause of the soup burning. On the one hand, it seems natural to say that it is. On the other, there seems to be no physical link: physically the door has nothing to do with the soup, and so counting the knock at the door as a cause seems odd.

If we sign up to the thesis of Dependence, we will count the knock at the door as a cause of the soup burning. But, Hall argues, combining this with Locality, Intrinsicness and Transitivity leaves us with problems.

Allowing cases of double prevention as causal seems to contradict Locality. It means that there can be causes which are not connected to their effects by intermediate events which are spatiotemporally contiguous; there can be gaps. Hall writes that even if action at a distance is possible, then it is an interesting and substantive fact about the world, not something which is trivial as a result of double prevention. One way out might be to allow omissions as part of the spatiotemporally contiguous chain. However, it is difficult to say where omissions are located. And even if we do locate omissions in the place where the event would have occurred if it had, the knock at the door does not seem spatiotemporally connected to the omission of stirring taking place in the soup on the stove.

To see the problem with the combination of Intrinsicness and Dependence,

we need to imagine the structure of events S which consists of e and all its causes back to some arbitrary time. Two structures of events with the same intrinsic character will have the same causal character, according to the Intrinsicness thesis. The pattern of dependence can, though, be changed by something extrinsic to the pattern of events which includes the burning soup and the events which led up to it. For example, if there had not been a knock at the door, I might have started to read a good book and become distracted so that I did not stir the soup anyway.

The problem with Transitivity is that it leads us to account events as causes when they don't make any difference to the final result. For example, I need to catch a train, but I have not left enough time to walk to the station, so I phone for a taxi. But the taxi driver gets lost on the way to the station, so I miss my train anyway. My missing my train depends on the taxi driver's getting lost, and that depends on my having phoned for a taxi in the first place. With Transitivity, that makes my calling a taxi a cause of my missing my train, but that seems wrong, because I would have missed my train anyway.

The upshot of all this we are in a double bind, according to Hall. If we start out with a counterfactual account of causation, we are going to have some problem cases to deal with. For some of these problem cases it seems like we need Locality, Intrinsicness and Transitivity. But in the case above, each of those three theses separately makes the case more puzzling. So we are pulled both ways about whether causation should involve these three theses as well as dependence.

10.4.4   Omissions

Hall also argues that Transitivity, Locality and Intrinsicness come into conflict with the thesis of Omissions: the claim that absences can be causes and effects. I discuss the issue of absence causation in detail in chapter 3. Here I will look specifically at Hall's approach to it.

Hall gives the following example: Suzy is flying a plane on a mission to bomb an enemy base. An enemy fighter ("Enemy") would have shot her down before she could complete her mission, but Billy shoots down Enemy before this can happen. So, Hall argues, the failure of Enemy to shoot down Suzy is clearly a cause of the success of Suzy's mission. But equally, it is not clear where that absence is located or if the absence of Enemy shooting down Suzy intersects with Suzy's actual flight. Essentially, when there is an absence involved in the causal chain then that seems to lead to spatio-temporal gaps, or at least to the possibility that such gaps exist.

Hall also argues that it must be the intrinsic character of a cause which makes it the cause of its effect. Absences, Hall writes, are extrinsic disabling factors which are not intrinsic to the causal chain leading up to an effect. So, he argues that we cannot both accept the intrinsicness of causes and the idea that absences can be causes.

Hall takes the view that causation is transitive, and that the transitivity of causation comes into conflict with absences being causes. Since Hall's example is rather complicated, however, I will attempt to describe the issue in more general terms. Suppose we have our effect e, our cause c, and our intermediate event d. If d occurs it will prevent e. Let's call the absence

of *d o*. So if absences can be causes then *o* is a cause of *e*. There is an intermediate stage *f*, between *c* and *d*. If *f* happens, it causes *d*, if nothing prevents it. *c* happens and causes *f*, but at the same time *c* also prevents *d*. *c* nearly indirectly causes, but also directly prevents *d*. Transitivity, then, means that *c* is a cause of *e*. Hall clearly feels this is the wrong answer, and clearly an absurdity, presumably since *c* makes no difference to *e* either way.

10.5 Hall's proposed solution: two concepts of causation

So we have five theses which are all plausibly true of causation, and which all seem to be needed in order to give the 'right' result in certain cases, but which also come into conflict. Hall's suggestion is that we find this conflict because there are two concepts of causation in play.

In one sense of causation, Dependence and Omissions are true; this is the dependence concept of causation. When we are talking about production on the other hand, Locality, Intrinsicness and Transitivity are true. Normally, when we think of a case of causation, the cause and effect are linked by both the ancestral of counterfactual dependence, and by production. However, cases of overdetermination are cases of production without dependence. And cases of double prevention are cases of dependence without production.

10.6   Hall's account of his two concepts

We have seen why Hall believes two concepts of causation are needed. In this section, I want to outline Hall's analysis of the two concepts. His analysis of dependence is very simple, his analysis of production less so. Hall's analysis of dependence is simply counterfactual dependence. This incorporates the theses of Dependence and Omissions. Hall's account of production is designed to incorporate the theses of Locality, Transitivity and Intrinsicness.

The analysis relies on the ideas of sufficiency and minimal sufficiency. Some set of events S occurs at t, and e (the effect) occurs at some later time t'. (This means no backwards in time causation, or simultaneous causation.) 'The laws', the premise that the events of S occur at t, and the premise that no other events occur at t, entail that e occurs at t'. This is what it is for S to be sufficient for e. Iff no proper subset of S is sufficient for e then S is minimally sufficient for e.

Hall would like to be able to simply say that the members of the unique minimally sufficient set for e at a time are the producers of e at that time. Unfortunately, there are producers that belong to no minimally sufficient set, and members of a minimally sufficient set which are non-producers. Hall provides examples in the form of neuron diagrams.

Hall's suggestion is that we deal with these by focusing on transitivity and the need for the producers to produce their effects using the right intermediaries. We need to imagine a timeline, from t to t'. e occurs at t'. At the earlier time, t, there is a unique minimally sufficient set for e, S. We can imagine that there are intervening times between t and t'; we can call these

$t0$ and $t1$. These two times also have their own minimally sufficient sets, $S0$ and $S1$. Hall writes:

(i) For each element of $S1$, there is at $t0$ a unique minimally sufficient set for it ...

(ii) the union of these minimally sufficient sets is $S0$

(Hall 2004, p. 265)

In other words, each member of the minimally sufficient set must itself have a minimally sufficient set (in the past). And, all of these sets bundled together must form the temporally first minimally sufficient set, $S0$. This is a way of making Hall's account apply to each step of the causal history of e.

Rather like Woodward's analysis of causation, Hall's analysis of production is not simple to state (no 'c causes e iff...') but involves several intermediate technical terms which have their own definitions. I will therefore avoid quoting his definition of production in full, but will set it out in stages.

- Production is 'the ancestral of **proximate causation**'.

- c is a **proximate cause** of e if both c and e and belong to a structure of events S. In simple cases, S will consist of e, and the **pure causal history** of e back to some earlier time. In harder cases, S needs to intrinsically match a nomologically possible structure of events S', which consists of an e duplicate, and the **pure causal history** of the e duplicate back to some earlier time.

- e occurs at t'. There is some earlier time, t. e's **pure causal history** back to some earlier time t is the structure of events consisting of the members of certain sets. These are the members of minimally sufficient sets for e at every time between t and t'. These sets must also meet the two conditions quoted above.

Explained crudely, producers of effects must be minimally sufficient for them, and also meet other criteria to include transitivity and intrinsicness. Hall's idea of production includes a modal notion of intrinsicness: two structures of events which match in intrinsic features will also match in causal features.

Hall's account of production is, then, a counterfactual account with extra attachments to ensure Locality, Transitivity and Intrinsicness.

10.7   Why these five theses?

The obvious question which springs to mind here is why Hall chooses to place the distinction in that particular place. We might question the need to group Dependence and Omissions together as one concept, and Locality, Intrinsicness and Transitivity together as the other and wonder if other combinations be possible which would satisfy all the requirements. Despite Hall's professed aim to see the big picture in this paper and avoid philosophical trench warfare, his argument is very focused around specific cases. It feels as though we need some other motivation for grouping the theses together in this way.

One reason for putting Dependence and Omissions together might be the method we have chosen for individuating absences. If we have chosen to un-

derstand each absence as the absence of a specific event in another possible world, then this fits well with a counterfactual analysis. A counterfactual analysis means we already committed to some theory of possible worlds (unless we take counterfactual claims as primitive) so once the concept is in place it might as well be put to work in helping us individuate absences too. Both thinking of causation as counterfactual dependence and thinking that absences can be causes show a commitment to a view whereby causation has to do with what might have happened and a turning away from a view of causation as something always involving actual particular physical events. This is not, however, an argument Hall makes.

There is also the question: why those five theses? Every single one of them has been doubted by some philosopher or other. Salmon (1998), Fair (1979) and Dowe (2000) all disagree with Dependence, since they offer an account which has nothing to do with counterfactuals. Similarly, Dowe denies Omissions. Meanwhile, Hall himself writes that Locality is only a contingent truth. Lewis's account does not involve Locality at all; nor does Woodward's (I discuss Woodward's interventionist account in chapter seven above). Intrinsicness would be denied by Woodward (2003) or Price et al. (2007) offering a non-reductive account of causation; they would not agree that whether a process is causal is determined by its intrinsic non-causal character, but would rather argue that the important question is whether the cause could be used to manipulate the effect. Adopting an interventionist or perspectivalist approach is in part to deny intrinsicness.

So in each case where Hall argues that our five theses do not go together, one option is simply to abandon one of the theses. However, Hall's five theses

are all plausibly true of causation. He argues convincingly that the conflicts between these theses explain problem cases in the philosophy of causation. Obviously one way of responding to these problem cases is to drop one of the theses about causation, but Hall offers an alternative to this approach, where we accept that all five theses are true, just not all true of the same kind of causation.

10.8   The wrong two concepts

At first, Hall's two concepts of causation look as though they are supposed to match up with a distinction which is popular in the philosophy of causation: between causation understood as something to do with counterfactual dependence, and causation understood as a physical process (Dowe 2009).

While few philosophers of causation adopt a fully fledged pluralism, it isn't uncommon to hold that there's a rough distinction between the two kinds of causation. In Lewis's case, it isn't pluralism because one of two isn't causation. In Dowe's case, it isn't pluralism because one of the two is a concept in our heads, not something which takes place in the world. But whatever the reasoning is, both put the divide in roughly the same place. It is odd, then, that Hall who is actually setting out to offer a pluralist view, puts the divide somewhere else.

So it seems natural to say something like this: there are two concepts, or two kinds of causation, or two ways of analysing it. On the one hand there is an account of causation as something to do with counterfactual dependence. This analysis of causation will be meant to be necessarily true and will allow

for things like omissions and absences. It will be the humanised face of causation; it will be the concept we use when we want to know whether causation is going on. As Woodward points out in his discussion of Salmon (1998), we don't tell a story about the causal process undergone by every relevant particle to tell a story about why something happened.

On the other hand, there is the biff account of causation. This analysis may or may not be contingent, but it will be something we had to find out about a posteriori, through our understanding of the laws of nature. It will be about processes and interactions, about energy and pushes and pulls. This will be the natural face of causation; it will describe processes which we wouldn't hesitate to say could go on in a world just like ours but without any minds. It will be the concept that can underwrite other causal claims with something concrete and physical.

The puzzling thing about Hall's two concepts of causation, then, is that they don't in fact match up with these two different definitions. Hall's definition of dependence is clearly a species of the first approach to causation described above. But his account of production is not a version of biff. Rather, it is a deductive account (more reminiscent of Hempel (1965) than anything else), with some additional criteria added, so that it will deal with counterexamples he finds troublesome.

Neither of Hall's two concepts reduce to the other. Dependence is not grounded by production; it is to be analysed in terms of counterfactuals. And production is not grounded by dependence, despite also being a counterfactual analysis in its own way.

This seems like an odd approach. To begin with, production, for Hall, is

meant to embody the three theses: Intrinsicness, Transitivity and Locality. It seems that some Dowe-like account which focuses on the conservation of energy or similar will do this naturally. We can imagine some idealised case like rock A striking rock B, somewhere in space. Of course the transfer of energy to rock B is intrinsic; it does not involve other objects, possible or actual, and transfer of energy is in itself non-causal. Of course the cause and effect have to be located in the same place; even for apparent cases at actions at a distance there has to be some kind of mediator such as a gravitational or electromagnetic field. Conservation of energy already has the right characteristics, prior to thought experiments showing us that these are features needed for causation.

Transference is however, not transitive. If A transfer some energy to B and B transfers some energy to C, it does not follow that A has transferred some energy to C. This is where it will make a difference what our favoured biff theory is. If we were to take Salmon's mark transmission theory, then it will be transitive, because it is the same mark being transferred. We can say that A transfers the mark to C by way of B. But if we take Dowe's theory where there's no real sense in which the same energy is transferred, we get a different result. If Mary pushes Jane and Jane pushes Emily, Mary has not pushed Emily by way of Jane, because it is not the same push. What is passed on is not a specific push but rather a conserved quantity of energy in the form of a push.

So not all biff-like accounts of causation will prefectly fit Hall's desiderata for production, but it is certainly possible to find one which does. Alternatively, we could take the ancestral of a conception of biff, to get something

similar which was transitive.

Dowe has another, related, worry about Hall's account. Dowe (2009, p. 37) points out that the two concepts of causation are not logically distinct. That is, Production under Hall's account entails Dependence. Absence causation is supposed to be a kind of Dependence without Production. But Dowe points out that 'had *a* obtained, *a* would have Produced *b*' implies 'not-*a* Dependence-caused not-*b*'. Dowe suggests that cases of Dependence supervene on possible-Production. His point is that the concepts of Dependence and Production are not sufficiently distinct, and therefore there is still a lack of logical independence between actual and possible causation.

This is another worry about Hall's account of causation: it lacks reductive potential. On the one hand, from the point of view of parsimony, positing two accounts of causation (both of which are presumably meant to be fundamental) is not helpful, since this multiplies the total number of fundamental relations in our ontology. Hall's response to the complaint that his account lacks parsimony is as follows. According to Hall, the principle of parsimony is basically like the suggestion that if you have lost your keys, you should start looking in the light part of the room, not the dark corners. So, he says that with causation it makes sense to start off looking for a univocal account. If that strategy fails, we should start looking in the dark corners ie searching for a pluralist account. In my chapter eight above, I argue for a similar view of parsimony: it is not a theoretical virtue for philosophical theories, in the sense that although it may be a useful tool when coming up with theories for the reason Hall gives, it does not make the theories any more likely to be true.

We might also be keen on reduction from the point of view of empiricism. We might be looking to reduce causation to something physical and empirically accessible, perhaps some relation found in fundamental physics. This would explain how it is possible that we get knowledge of causation in the first place. Hall's account is unhelpful here, since neither of his accounts involve any reduction to anything biff-like. Even his account of production is counterfactual, albeit in a different way from his account of dependence.

10.9   Conclusion

My view is that Hall is right to suggest two concepts of causation. I agree that one of these should be counterfactual and be such that some absences can be counted as causes and effects. The other account should involve locality and intrinsicness. I find the idea that there are two concepts convincing; it explains why we are pulled two ways when considering some cases, and that our responses change depending on how we think of causation.

However, I don't entirely agree with Hall about exactly what the two concepts should be. The production account of causation should, I think, be based on a process theory such as Dowe's; if the account is based on physical interaction it explains why locality and intrinsicness should hold. Finally, I think that while there are two kinds of causation, we should consider the possibility that one could ground another, thus giving the account as a whole account more reductive potential. I explore the possibility of combining two concepts of causation into one account in the next chapter.

Chapter 11

A single account of the two kinds of causation

11.1   Introduction

In the previous chapter I have argued that there are two kinds of causation. In this chapter I will suggest the following:

1. One of the two kinds of causation is more fundamental than the other.

2. The more fundamental kind of causation grounds the less fundamental kind.

I will not offer conclusive arguments for either of these claims, but I will suggest that taking this approach would be in line with many of our intuitions about causation. I will also deal with a couple of potential objections to show that the approach could work in principle.

I am suggesting that facts about one kind of causation ground facts about the other, but that providing an analysis of the less fundamental kind of causation in terms of the more fundamental kind will not be a simple task. I will, however, deal with the obvious objections to such an analysis: firstly, that it would be disjunctive, and secondly that it does not make sense to have causation by absences taking place at one level and not at the other.

Firstly, I will summarise an account due to Strevens (2013) and his argument that while there are two kinds of causation, one is more fundamental than the other. Next, I will argue that the less fundamental kind of causation is grounded in the more fundamental kind. I will argue that Strevens is basically right when he says that facts about causal difference making stand to facts about causal influence as facts about centres of mass stand to mass. Finally, I will defend the view against worries about disjunctiveness and absence causation.

A note on terminology: Strevens uses the term causal difference making to denote the less fundamental kind of causation, and causal influence to denote the more fundamental kind of causation. I will borrow this terminology, without committing myself to giving exactly the same characterisations of the two kinds of causation as Strevens. This distinction comes very close to aligning with the distinction between causation and biff which Lewis (2004b) makes; however, I will use Strevens's terms as I believe that both kinds are causation of a sort[1]. I do share Strevens's view that Dowe's account of causation provides a good candidate for our analysis of causal influence. I

---

[1]I do not share Lewis's intuition that causation may exist in any possible world, which is his motivation for calling 'biff' something other than 'causation' (Lewis 2004b).

am less inclined to accept Strevens's deductive account of causal difference making, but I do share his view that the second less fundamental kind of causation is essentially about making a difference.

11.2 Strevens's account of causation

In this section I will explain the details of Strevens's account of the two kinds of causation (Strevens 2013), and briefly compare his account with the account due to Hall (2004).

Strevens offers a re-unified account of causation largely as a reply to Hall. He writes that Hall is correct that there are two concepts of causation, but argues that this need not amount to pluralism since the two account share a 'fundamental-level causal reality' (Strevens 2013, p. 308). In this section I will give an outline of Strevens' account. As we will see, he differs from Hall in some other ways as well as on the issue of pluralism, in the details of of his analyses. This is what makes the re-unified account possible.

The first kind of causation, according to Strevens, is causal influence. We can understand the idea of causal influence as synonymous with physical causation, or biff. Both are ways of describing cases of one particular directly and concretely causing or influencing another: of picking out a singularist, physical form of causal relation. In fact, Strevens (2013, p. 14) invites us to swap in our favourite physical account of causation here, for example Dowe's conserved quantity theory (Dowe 2000). This makes up the more fundamental causal reality. Any causal facts that are true are true in virtue of facts about causal influence.

The other kind of causation is causal difference making. Roughly speaking, for c to be a cause of e, c needs to be part of the web of causal influences leading up to e, and also make a difference to whether e happens or not. Causal difference making could be cashed out in terms of counterfactuals, but Strevens prefers what he calls the kairetic account.

According the kairetic account, we can find the causal difference maker for e in the following way. We take a chunk of the web of causal influence which leads up to e, and make it the premise of a deductive argument; the conclusion of the argument is that e occurs. This "chunk" need not be all of e's causal history. The deductive argument will contain initial conditions and boundary conditions to specify how far in space and how far back in time this "chunk" extends. The premises of the argument, Strevens tells us, must include the laws of nature which make it possible for the causal influences to influence e, as well as the details of the influences which are jointly sufficient to bring e about.

We then abstract as much as possible without invalidating the causal implication. This will give us the minimally sufficient difference makers for that chunk of the web. (To get all the difference makers we would have to apply this to every part of the web.) For a statement to be more abstract than another statement it must, according to Strevens, (1) be implied by that statement and (2) has the same subject matter as that statement, or a subset of the same subject matter. The abstraction process should continue until the argument is either non-valid, or is non-causal. An argument is non-causal if it is purely analytically valid, for example, 'the window smashed therefore the window smashed' is logically valid but is non-causal. The abstraction

must not be disjunctive.

For example, suppose that the event e is the window's smashing. The web of causal influence leading up to e will include the rock thrown at the window, but also the light shining on the window, the wind and air pressure, sounds, and all other events and conditions happening in the spatiotemporal region before the window shatters. Obviously this is too much, so Strevens says a 'chunk' of the web: presumably, just the events nearby and immediately before. The statement that all this occurs, along with the laws of nature, becomes the premise of a deductive argument, with the conclusion being that the window shatters. We then make the argument as abstract as possible, without invalidating the causal implication. So for example, we can take the statement that each of the atoms on the front surface of rock hits the window, and abstract it to simply saying that the rock hits the window. We can abstract the statements about each molecule in the air which hits the window, and say that the air pressure was such-and-such. Whatever statements are left at the end of this process are the minimally sufficient difference makers for the window's smashing, in this case, the rock's hitting the window, as well as certain background conditions such as the air pressure.

11.2.1 A comparison of Hall's and Strevens's accounts

Hall (2004) and Strevens (2013) actually have quite different views on the nature of the two kinds of causation. Recall that Hall's two accounts take the form of counterfactual dependence on the one hand, and production on the other. The analysis of production is complex, but Hall argues that the best

approach is to take a basic account, such as counterfactual dependence, as our blueprint. We imagine a system of events which count as a cause because counterfactual dependence holds between them and the effect. Then, very roughly, a cause will be the system of events which resembles in its intrinsic features that system where counterfactual dependence takes place.

The result is that we have taken counterfactual dependence to be our basic causal relation, and we analyse production by elaborating on and modifying that. What is notably missing here is any notion of biff. It sounds at first glance as though production is meant to be something of this sort, but in fact it is nothing of the kind. It would be more accurate to say that Hall offers the simple counterfactual account and the elaborate counterfactual account.

Strevens, on the other hand, offers two accounts, one of which is essentially biff, and the other of which is deductive, as I outline above. It is odd then, that Strevens writes that he agrees with the 'important' (Strevens 2013, p. 309) aspects of Hall's account, which seems to imply the substantive content of the accounts, when in fact Hall and Strevens do not have the same view about what the two kinds or concepts are. Strevens puts the divide in a much more usually accepted place than Hall.

11.3   Causal structure and causal metaphysics

In this section I will briefly describe a two tier account of causation due to Handfield et al. (2008). Their account provides a contrast to Strevens's and shows that there is more than one possible way of cashing out the details of how two accounts of causation might be linked together.

Handfield et al. (2008) offer a different approach to integrating two concepts of causation. Their thought is that causal modelling gives us the structure of causation, whilst causal process theories of causation such as Dowe's give us the metaphysics. Handfield et al. (2008) also argue that the relationship between these two levels is one of grounding.

According to their account, every arc of a model must correspond to a salient possible connecting process. An arc of a model connects variables which have values. The values of these variables are grounded in events. So for example when Suzy is throwing a rock at a bottle and it smashes, we have the variables *Suzy throws* and *bottle smashes*, which can have the value of true or false. If *Suzy throws* is true, for example, then that is grounded by the event of Suzy throwing the rock. If there is a causal process connecting the grounding events for the variable, then there is a process corresponding to the arc which connects those variables. The grounding relation between the variables and the events in question is truthmaking[2]. An event's occurring makes it true that a variable has a certain value. The authors suggest we should look to Dowe or Salmon for a theory of what a causal process is, and that for a process to connect two events is for those events to be constituents of it.

The authors have their own approach to accounting for causation by and of absences, and making sense of the way causal judgements can sometimes depend on perspective. I will discuss this in a section below.

---

[2]One understanding of truthmaking is that it is a kind of grounding (Fine 2012, p. 8).

11.4   One kind of causation grounds the other

In this section I will suggest that Strevens is right when he describes to rela-
tionship of dependence between causal difference making and causal influence
as similar to that between centres of mass and mass.

Strevens writes that any fact about causal difference making can be ex-
pressed as facts about causal influence. He compares it to facts about centres
of mass, and facts about mass. The former are useful, and facts about causal
difference making are also very useful to us. But if we could say everything
there is to say about mass, without adding facts about centres of mass. This
is what Strevens believes makes his account of causality unified. There is just
one 'fundamental-level causal reality' (Strevens 2013, p. 308), and causal dif-
ference making is a different, useful, but ultimately ideologically lightweight
way of talking about it.

Although Strevens does not use the word 'grounding' to explain the rela-
tionship between causal influence and causal difference making, it seems as
though the concept he has in mind is something similar. I discuss the notion
of grounding in chapter one; it is a relation whereby something less funda-
mental depends non-causally on something more fundamental. In chapter
one I suggest that if we reductively analyse causation in terms of a, then we
are saying that facts about a ground facts about causation. For short, we
could say that a grounds causation. I want to add, though, that we could
say that one thing grounds another, or more properly that facts about the
more fundamental concept ground facts about the less fundamental concept,
without offering an analysis. An analysis would be a bonus, as it would allow

us to say just how the more fundamental grounds the less fundamental. But we can still say that grounding is the relation between the facts involved without saying how it relates them.

I think something like this is going on here. Causal influence grounds causal difference making, or to be more precise, facts about causal difference making ground facts about causal influence. Strevens offers an account in his paper of how this works, showing that in principle we could analyse causal difference making in terms of causal influence.

Strevens's analogy with mass and centres of mass is helpful. Facts about centres of mass supervene on facts about mass. There can be no difference to facts about centres of mass without a difference to facts about mass, and not vice versa. But it is also interesting to ask what the relationship is between mass (itself, not facts about it), and centres of mass. For one thing, it is clear that mass is more fundamental; gravity is one the of the four fundamental forces in the universe, while being a centre of mass is not. For another, it is clear that something's being a centre of mass of some other thing depends on that other thing's having mass. This dependence is non-causal, so this looks like a classic case of grounding. Finally, if we were to give a reductive definition of a centre of mass, the main conceptual ingredient in the analysis would be mass. To sum up, mass grounds centres of mass. Being a centre of mass reduces to having mass. Mass is fundamental and centres of mass depend upon it.

In his influential paper 'On What Grounds What', Schaffer (2009, p. 378) argues that one way of looking at the grounding relation would be to see it as abstraction. The derivative entities must, he writes, be 'latent within the

substances'; that is, the non-fundamental things must be already present in some way in the fundamental things. Abstraction brings out certain aspects and changes our focus.

Strevens describes that process of getting from facts about causal influence to fact about causal difference making as part one of abstraction. Causal difference making is a more abstract way of looking at the same causal reality (Strevens 2013, p. 34).

This then leaves us with a metaphysics of causation which has a two-tier structure. On the bottom layer we have fundamental non-causal facts. These ground the next layer, which is facts about causal influence. These in turn ground the top layer, which is facts about causal difference making.

An analysis, if found, would tell us how we get from one layer to the one above it. However, that analysis might not be of the form 'c makes a causal difference to e if and only if...'. That form of analysis assumes that the relata are the same at each level. So for example if we think causation takes place between events (that events are the causal relata) then in analysing causation we would traditionally look for some other relation(s) which held between events, which could plausibly be what grounds the causal relation. But at each stage here I would argue that we are looking at relations between different kinds of relata. At the level of causal influence we are looking at particulars, specifically objects which can have worldlines and can possess a conserved quantity. At the level of causal difference making we are looking at facts, composed of a particular, a property and the relation of either instantiation or anti-instantiation (for negative facts, see chapter 3 above).

11.5 Defending the idea of a two-tier analysis

In this section I will look at some possible worries for a two tier account of causation: absence causation, and the worry that the account would end up being disjunctive.

11.5.1 Absence causation

Causation by and of absences is a persistent issue for any account of causation. However, the problem presents itself differently for each account of causation. In the case of a two-tier account of causation, the problem is that absence causation should be possible at the less fundamental level, but not at the more fundamental level. So the obvious question is how absence causation can be possible at one level but not at another.

There is no absence causation at the level of causal influence. Causation understood as one particular pushing on another, or as transference energy or momentum between objects, leaves no room for absence causation. At the level of causal difference making, absence causation makes more sense. It is easy to see how an absence can make a difference to something, or how a difference can be made to an absence.

However, I will argue that there is nothing metaphysically puzzling about absences being involved as relata on one level and not on another. Take an example of another relation, such as spatial contiguity. My hand is on the desk; they are clearly spatially contiguous. Looked at on another level, however, both my hand and the surface of the desk are made up of atoms which are mainly empty space. Although the electrons in my hands and

the electrons in the desk are interacting to stop me from pushing my hand through the desk, few if any of the sub-atomic particles in my hand are actually coming into contact with the sub-atomic particles which form the surface of the desk. On one level, the relation of contiguity between my hand and the desk is definitely between positive, actual particulars. On another level, however, the relation is between areas of empty space between electrons: absences. Relations play out differently depending on whether we are looking on a fundamental or a less fundamental level.

More specifically, I think that at the level of causal influence there is no causation by and of absence, since causation consists of causal processes and causal interactions. The processes and interactions are between particular objects. There cannot be negative objects. But at the level of causal difference making I have argued that causation is between facts, and I have argued that negative facts are not problematic. But the facts which are the causal relata at the causal difference making level are composed of particulars and properties. In other words, causation between facts is not the most fundamental kind of causation and so of course it has non-fundamental relata, facts.

Strevens's approach to absences is in two parts. Firstly, he writes that at the level of causal difference making things that part of the causal web can certainly have negative properties. This ensures that absences can be causes; for example I can have the property of not watering the plants. That still leaves us with the problem of how to avoid too many absence causes; for example, the Queen also has the property of not watering the plants. But, if they are my plants, we don't want to cite her not watering them as a cause of

their dying. Strevens's answer to this is what he calls frameworked difference making. Briefly, the causal framework is a set of propositions concerning the causal web which are held to be true by those who are making assertions about causal difference making. This framework changes our process of abstraction so that only the right absences come out as causes. (Strevens 2013, p. 313).

Handfield et al. (2008) have a somewhat different approach to dealing with absence causation. For them, similarly, absence causation can take place at the level of causal difference making, but not at the level of causal influence. Their approach is to say that the negative facts which are causes and effects are grounded by actual events. The actual event in question is the event which happened instead of the event which did not occur. For example, the fact that I did not stir the soup is grounded by the actual concrete event of my answering the door. They argue that, though this actual event may not be a truthmaker for the negative fact, it is nonetheless physically incompatible with it. (Handfield et al. 2008, p. 158)

So accounts are available which detail how absence causation could be possible on one level, but not on the more fundamental level which grounds it.

11.5.2 Against worries about disjunction

In this section I will defend the idea that we can have one more fundamental kind of causation grounding the other less fundamental kind against the objection that this will lead the analysis of the less fundamental kind of

causation being disjunctive.

The main worries that we need to consider relating to unified accounts are the same as the worries which Lewis considers reasons to not adopt the Canberra Plan when analysing causation. Lewis (2004b) rejects the idea that causation could supervene on biff-relations, because of the worry that something else could fill the causation role, which in our world is filled by biff, in some other possible world. The thought is that our analysis of causation should be necessarily true; we should say what causation is in every possible world. An analysis involving biff might not do this if biff relations do not exist in some other possible world, that is, if it is possible for the laws of nature involving the transfer of conserved quantities to be otherwise.

The kind of two-tier account I am endorsing here could fall prey to the same problem. In this case the concern would be that although causal influence grounds causal difference making, something else might have grounded causal difference making, or does ground it in another world.

In chapter six I argue that the answer is to claim that whatever underwrites causation underwrites it necessarily. To be more specific, in this case facts about causal influence necessarily ground facts about causal difference making. The idea is that this is an a posteriori necessity. This is suggestion is due to Handfield et al. (2008, p. 156). This means that other possible worlds in which causation takes place, must also have laws of nature enough like ours for causal influence to work in the same way. Worlds which have laws of nature which are different enough from ours to break Dowe's account of causation will not count as having causation at all. There may be something which happens there which is a bit like causation, but it isn't causation. This

answer seems like a good one when we consider that we don't know what a world with different laws governing mass-energy and charge (instead of conservation laws) would even look like. A world with laws of nature different from ours at such a basic level does seem unrecognisable and it doesn't seem implausible to think that if we could see what it was like, we wouldn't recognise anything like causation in it.

11.6 Conclusion

Strevens (2013) argues that, whilst there are two kinds of causation, causal influence and causal difference making, there is only one 'fundamental-level causal reality' (Strevens 2013, p. 308). We can expand on this idea to suggest that facts about causal influence ground facts about causal difference making. I have argued that this has the potential to work in principle; apparent problems with absences causation and disjunction are surmountable. Strevens account offers one approach for explaining how the two accounts of causation might fit together; there is another example due to Handfield et al. (2008).

Chapter 12

Conclusion

In this concluding chapter I will briefly describe my conclusions from each part of my thesis: the causal relation, reductive accounts of causation, defending reductive analysis, and two kinds of causation.

12.1   The Causal Relation

When we analyse causation, we are saying that the analysis relation obtains between the analysandum (causation) and the analysans. I have argued (in chapter one) that this relation is one of grounding, specifically, that the facts about the analysans ground the facts about the analysandum. To give an analysis of causation, then, is to make a claim about what grounds causal facts.

The causal relata what the causal relation holds between: the cause and the effect. I have argued (in chapter two) that causation, at least at the level of causal difference making, is best understood as between facts (or states of

259

affairs).

There are many clear cases of causation where one of the relata is negative, or is an absence. These cases cause a variety of problems. I have argued (in chapter three) that absence causation is possible at the level of causal difference making, where the absences can be understood as negative facts, but not at the level of causal influence.

12.2   Reductive accounts of causation

I have described three approaches to reductively analysing causation: Lewis's (chapter four), Dowe's (chapter five) and the Canberra Plan (chapter six). I have argued that Dowe's account of causation is convincing, but only as one kind of causation, as causal influence. I have discussed the worry that a Canberra Plan approach to analysing causation will lead to a disjunctive analysis, and concluded that this is not a problem as long as we hold that whatever grounds causal facts does so necessarily.

12.3   Defending reductive analysis

I have discussed Woodward's non-reductive account of causation, in chapter seven. I have argued that his motivation for providing a non-reductive account is not compelling. I then discuss two possible reasons for seeking a reductive account: ideological parsimony (chapter eight) and empiricism (chapter nine). I argue that ideological parsimony does not give us a good motivation for reductive analysis, but empiricism does.

12.4 Two kinds of causation

Finally, I have argued in chapters ten and eleven that there are two kinds of causation: causal influence and causal difference making. Causal influence grounds causal difference making.

Bibliography

Anscombe, G. E. M. (1971). *Causality and Determinism*. Cambridge University Press.

— (1993). "Causality and Determination". In: *Causation*. Ed. by E. Sosa M. Tooley. Oxford Up, pp. 88–104.

Armstrong, D. M. (1989). *A Combinatorial Theory of Possibility*. Cambridge University Press.

— (1997). "Singular Causation and Laws of Nature". In: *The Cosmos of Science*. Ed. by John Earman and John Norton. University of Pittsburgh Press, pp. 498–511.

— (2004). *Truth and Truthmakers*. Cambridge University Press.

Baumgartner, Michael (2008). "Regularity Theories Reassessed". In: *Philosophia* 36.3, pp. 327–354.

Beall, JC (2000). "On Truthmakers for Negative Truths". In: *Australasian Journal of Philosophy* 78.2, pp. 264–268.

Beebee, Helen (2006). *Hume on Causation*. Routledge.

Cartwright, Nancy (2004). "Causation: One Word, Many Things". In: *Philosophy of Science* 71.5, pp. 805–819.

Chalmers, David J. (1996). *The Conscious Mind: In Search of a Fundamental Theory.* Oxford University Press.

Collingwood, R. G. (1998). *An Essay on Metaphysics.* Oxford University Press.

Davidson, Donald (1980). *Essays on Actions and Events.* Oxford University Press.

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

— (2009). "Absences, Possible Causation, and the Problem of Non-Locality". In: *The Monist* 92.1, pp. 23–40.

Ducasse, Curt John (1969). *Causation and the Types of Necessity.* New York, Dover Publications.

Duhem, Pierre Maurice Marie (1954). *The Aim and Structure of Physical Theory.* Princeton, Princeton University Press.

Emmet, Dorothy (1984). *The Effectiveness of Causes.* Macmill.

Fair, David (1979). "Causation and the Flow of Energy". In: *Erkenntnis* 14.3, pp. 219–250.

Fine, Kit (2012). "Guide to Ground". In: *Metaphysical Grounding.* Ed. by Fabrice Correia and Benjamin Schnieder. Cambridge University Press, pp. 37–80.

Frege, Gottlob (1879). *Begriffsschrift: Eine Der Arithmetische Nachgebildete Formelsprache des Reinen Denkens.* L. Nebert.

Frisch, Mathias (2005). *Inconsistency, Asymmetry, and Non-Locality: A Philosophical Investigation of Classical Electrodynamics.* Oxford University Press.

Gasking, Douglas (1955). "Causation and Recipes". In: *Mind* 64.256, pp. 479–487.

Glynn, Luke (2013). "Causal Foundationalism, Physical Causation, and Difference-Making". In: *Synthese* 190.6, pp. 1017–1037.

Hall, Ned (2004). "Causation and Counterfactuals". In: ed. by Ned Hall, L. A. Paul, and John Collins. Cambridge, Mass.: MIT Press. Chap. Two Concepts of Causation.

Handfield, Toby, Charles Twardy, Kevin Korb, and Graham Oppy (2008). "The Metaphysics of Causal Models: Where's the Biff?" In: *Erkenntnis* 68.2, pp. 149–68.

Hempel, Carl Gustav (1965). "Aspects of Scientific Explanation". In: *Philosophy and Phenomenological Research*. Free Press, p. 504.

Hesslow, Germund (1976). "Two Notes on the Probabilistic Approach to Causality". In: *Philosophy of Science* 43.2, pp. 290–292.

Hitchcock, Christopher (2001). "The Intransitivity of Causation Revealed in Equations and Graphs". In: *Journal of Philosophy* 98.6, pp. 273–299.

— (2007). "Three Concepts of Causation". In: *Philosophy Compass* 2.3, pp. 508–516.

Huemer, Michael (2009). "When is Parsimony a Virtue?" In: *Philosophical Quarterly* 59.235, pp. 216–236.

Hume, David (2000 [1739]). *A Treatise of Human Nature*. Oxford University Press.

Illari, Phyllis and Federica Russo (2014). *Causality: Philosophical Theory Meets Scientific Practice*. OUP Oxford.

Jago, Mark and Stephen Barker (2012). "Being Positive About Negative Facts". In: *Philosophy and Phenomenological Research* 85.1, pp. 117–138.

Kim, Jaegwon (1973). "Causation, Nomic Subsumption, and the Concept of Event". In: *Journal of Philosophy* 70.8, pp. 217–236.

— (2008). "Reduction and Reductive Explanation : Is One Possible Without the Other?" In: *Being Reduced: New Essays on Reduction, Explanation, and Causation.* Ed. by Jakob Hohwy and Jesper Kallestrup. Oxford University Press.

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

Kvart, Igal (1991). "Transitivity and Preemption of Causal Relevance". In: *Philosophical Studies* 64.2, pp. 125–160.

Lewis, David (1970). "How to Define Theoretical Terms". In: *Journal of Philosophy* 67.13, pp. 427–446.

— (1973a). "Causation". In: *Journal of Philosophy* 70.17, pp. 556–567.

— (1973b). *Counterfactuals.* Blackwell Publishers.

— (1979). "Counterfactual Dependence and Time's Arrow". In: *Noûs* 13.4, pp. 455–476.

— (1986). *On the Plurality of Worlds.* Blackwell Publishers.

— (2004a). "Causation as Influence". In: *Causation and Counterfactuals.* Ed. by John Collins, Ned Hall, and L. A. Paul. MIT Press, PAGE REFS.

— (2004b). "Void and Object". In: *Causation and Counterfactuals.* Ed. by John Collins, Ned Hall, and L. A. Paul. MIT Press, pp. 277–290.

Liebesman, David (2011). "Causation and the Canberra Plan". In: *Pacific Philosophical Quarterly* 92.2, pp. 232–242.

McDermott, Michael (1995). "Redundant Causation". In: *British Journal for the Philosophy of Science* 46.4, pp. 523–544.

Mellor, David (1995). *The Facts of Causation.* Routledge.

— (2004). "For Facts as Causes and Effects". In: *Causation and Counterfactuals.* Ed. by Ned Hall, L. A. Paul, and John Collins. Cambridge, Mass.: MIT Press, pp. 309–23.

Menzies, Peter (1996). "Probabilistic Causation and the Pre-Emption Problem". In: *Mind* 105.417, pp. 85–117.

Menzies, Peter and Huw Price (1993). "Causation as a Secondary Quality". In: *British Journal for the Philosophy of Science* 44.2, pp. 187–203.

Molnar, George (2000). "Truthmakers for Negative Truths". In: *Australasian Journal of Philosophy* 78.1, pp. 72–86.

Nolan, Daniel (1997). "Quantitative Parsimony". In: *British Journal for the Philosophy of Science* 48.3, pp. 329–343.

Nozick, Robert (1969). "Newcomb's Problem and Two Principles of Choice". In: *Essays in Honor of Carl G. Hempel.* Ed. by Nicholas Rescher. Reidel, pp. 114–146.

O'Connor, Timothy (1996). "Why Agent Causation?" In: *Philosophical Topics* 24.2, pp. 143–58.

O'Leary-Hawthorne, John and Andrew Cortens (1995). "Towards Ontological Nihilism". In: *Philosophical Studies* 79.2, pp. 143–165.

Paul, L. A. (2000). "Aspect Causation". In: *Journal of Philosophy* 97.4, pp. 235–256.

Paul, L. A. and Ned Hall (2013). *Causation: A User's Guide.* Oxford.

Pereboom, Derk (2004). "Is Our Conception of Agent-Causation Coherent?" In: *Philosophical Topics* 32.1/2, pp. 275–286.

Plantinga, Alvin (1976). "Actualism and Possible Worlds". In: *Theoria* 42.1-3, pp. 139–160.

Price, Huw and Richard Corry (2007). *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford University Press.

Psillos, Stathis (2002). *Causation and Explanation*. Routledge.

Quine, Willard V. O. (1951). "Two Dogmas of Empiricism". In: *Philosophical Review* 60.1, pp. 20–43.

— (1963). "On Simple Theories of a Complex World". In: *Synthese* 15.1, pp. 103–106.

Ramsey, F. P. (1990). *F.P. Ramsey: Philosophical Papers*. Cambridge: Cambridge University Press.

Rosen, Gideon (1990). "Modal Fictionalism". In: *Mind* 99.395, pp. 327–354.

— (2010). "Metaphysical Dependence: Grounding and Reduction". In: *Modality: Metaphysics, Logic, and Epistemology*. Ed. by Bob Hale and Aviv Hoffmann. Oxford University Press, pp. 109–36.

Russell, Bertrand (1912). "On the Notion of Cause". In: *Proceedings of the Aristotelian Society* 7, pp. 1–26.

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

— (1998). *Causality and Explanation*. Oxford University Press.

Schaffer, Jonathan (2000). "Trumping Preemption". In: *Journal of Philosophy* 97.4, pp. 165–181.

— (2005). "Contrastive Causation". In: *Philosophical Review* 114.3, pp. 327–358.

— (2009). "On What Grounds What". In: *Metametaphysics: New Essays on the Foundations of Ontology*. Ed. by David Manley, David J. Chalmers, and Ryan Wasserman. Oxford University Press, pp. 347–383.

— (2014). "What Not to Multiply Without Necessity". In: *Australasian Journal of Philosophy* 93.4, pp. 644–664.

Sober, Elliott (1984). "Two Concepts of Cause". In: *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1984, pp. 405–424.

Stalnaker, Robert C. (1976). "Possible Worlds". In: *Noûs* 10.1, pp. 65–75.

Steglich-Petersen, Asbjørn (2012). "Against the Contrastive Account of Singular Causation". In: *British Journal for the Philosophy of Science* 63.1, pp. 115–143.

Steward, Helen (2012). *A Metaphysics for Freedom*. OUP Oxford.

Strawson, Peter F. (1985). "Causation and Explanation". In: *Essays on Davidson: Actions and Events*. Ed. by Bruce Vermazen and Merrill B. Hintikka. Oxford University Press, pp. 115–35.

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

— (2013). "Causality Reunified". In: *Erkenntnis* 78.2, pp. 299–320.

Tallant, Jonathan (2010). "Not a Total Failure". In: *Philosophia* 38.4, pp. 795–810.

Von Wright, G. H. (1971). *Explanation and Understanding*. Cornell University Press.

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