Counterfactuals and Counterparts: Defending a neo-Humean theory of causation

By

Neil McDonnell MA (Hons) MLitt Glas

A thesis submitted to Macquarie University
and to The University of Glasgow
in fulfilment of the requirements for the degree of
Doctor of Philosophy
Department of Philosophy (Macquarie)
School of Humanities, College of Arts (Glasgow)
February 2015

Typeset in LaTeX 2_ε.

Except where acknowledged in the customary manner, the material presented in this thesis is, to the best of my knowledge, original and has not been submitted in whole or part for a degree in any university.

Neil McDonnell MA (Hons) MLitt Glas

Acknowledgements

I dedicate this thesis to my mini-clan: Sarah, Finn and Aoife who fuelled this thesis with hugs, smiles and incredible patience.

I am grateful to the Arts and Humanities Research Council, Macquarie University and the Soluis Group for generous financial assistance. I thank the Bank of Mum and Dad for their lenient terms and highly competitive rates.

I could not have completed this thesis without the supererogatory efforts of Stephan Leuenberger and Peter Menzies who were my primary supervisors in Glasgow and Macquarie respectively. Martin Smith at Glasgow, my adjunct supervisor, also gave extremely helpful comments and advice. Albert Atkin at Macquarie fought the good fight on my behalf. I am considerably indebted to them all.

My colleagues at Glasgow are fantastic and their support has been immeasurable. Special mention is due to Umut Baysan, Michael Brady, Ben Colburn, Jen Corns, Dudley Knowles, Fraser MacBride, and Fiona Macpherson in this regard. In the broader philosophical community I have had difference-making help and advice from Racheal Briggs, Alan Hájek, Daniel Nolan, Jonathan Schaffer and Brad Weslake.

Most importantly, I have had the support of my parents Declan and Elizabeth, and my brother Martin.

So, in short: I would like to thank my friends for keeping me sane, my colleagues at Glasgow for keeping me in good company, my supervisors for keeping me straighter and narrower than I could have ever hoped to be on my own and my family most of all for simply keeping me.

Abstract

Whether there exist causal relations between guns firing and people dying, between pedals pressed and cars accelerating, or between carbon dioxide emissions and global warming, is typically taken to be a mind-independent, objective, matter of fact. However, recent contributions to the literature on causation, in particular theories of contrastive causation and causal modelling, have undermined this central causal platitude by relativising causal facts to models or to interests. This thesis flies against the prevailing wind by arguing that we must pay greater attention to which elements of our causal *talk* vary with context and which elements track genuine features of the world around us. I will argue that once these elements are teased apart we will be in a position to better understand some of the most persistent problems in the philosophy of causation: pre-emption cases, absence causation, failures of transitivity and overdetermination. The result is a naturalist account of causation, concordant with the contextual variability we find in our ordinary causal talk, and parsimonious with respect to the theoretical entities posited.

Contents

1

Introduction

What is it for one thing to cause another? The answer to this question will have an impact on questions of responsibility and agency, on questions of the interaction between the body and the mind and, most obviously, it will impact on causal theories of knowledge and perception. And yet, whilst an answer would have wide import, there is currently no philosophical consensus as to what that answer might be.

One might think that we demonstrate a mastery of the causal concept in our ability to manipulate the world around us. After all, if we didn't know what caused what, then we wouldn't be able to drive a car, operate a television or take a photograph. Plainly we already understand many cases of cause and effect! However the question is not 'what are the causes?' but rather 'what is it about the causes that make them causes?' and this is a notion we are yet to master.

The recent literature in the philosophy of causation has produced a wide variety of views which in turn have produced a plethora of problematic cases that appear to refute them. Reading this literature, one can detect a growing pessimism about the prospects of our ever being able to give a really informative and satisfying account of causation, one that conforms to our ordinary understanding and identifies some unifying feature that all cases of causation have in common. In particular there is great pessimism about our being able to give an account which captures the idea that causation is a *natural* rather than mind-dependent feature of the world around us.

In this thesis I will swim against that pessimistic tide. I will argue that we can understand causation in terms of counterfactual dependence relations that hold between events in our world, and that we can explain and defuse the most problematic counter-examples that have arisen to date.

1.1 Overall Aims

I take it that an adequate account of causation must be able to say what someone means when they say 'c caused e'. So, my first aim will be:

> (I) To account for our everyday causal assertions.

That first aim concerns our causal talk, not the metaphysics of causation itself. Of the causal relation Strawson said:

> We sometimes presume, or are said to presume, that causality is a natural relation which holds in the natural world between particular events or circumstances, just as the relation of temporal succession does, or that of spatial proximity. [Strawson, 1992, p.109]

Like temporal succession or spatial proximity, the relation of causation, we presume, holds or does not hold between c and e mind-independently. To which Menzies adds:

> If causation is a natural relation between events, then this relation should hold no matter how the events are described. In other words, causal propositions should be extensional in the sense that the substitution of one event nominal for a second co-referential event nominal in a causal proposition should preserve the truth-value of the proposition. [Menzies, 2009, p.346]

Menzies takes it to be our central causal platitude that causation is a natural, mind-independent relation. I hasten to add that Menzies identifies this notion in order to refute it. That refutation hinges, however, on there being no viable theory of causation which can conform to this platitude and account for our causal talk. So, in addition to the first aim of this thesis regarding causal talk, I add the following aim regarding the metaphysics of causation:

> (II) To give an account of the mind-independent, objective standard for causal connectedness between events.

In other words I will hypothesise that there is such a relation and proceed to see if there is a viable theory that can accommodate that hypothesis.

What connects (I) and (II)? The first concerns what we say and the second what the world is like, so insofar as they pertain to the same topic we might hope that that the account of causation that is offered meets both requirements under a single account. That does not require that our causal talk be analysed in one way and that our metaphysics be analysed in the very same way, but any strategy that aims to offer two different analyses owes an explanation as to how they are connected. So my third aim will be:

> (III) To explain the relation of (I) and (II).

I will aim to meet these requirements by advocating a two-tier account of causation. The first tier will concern the truth conditions for our causal talk and the second tier will concern the conditions under which causation takes place in the world. Both will be subject to a counterfactual analysis which stems from, but is not identical to, David Lewis's 1973 analysis of causation. Regarding such a two-stage strategy within the counterfactual tradition, Menzies says:

> [I]t would surely be a surprising fact, requiring elaborate explanation, if our framework for conceptualizing causation used in two different but crucial ways the very same idea of difference-making.

I will attempt to meet this explanatory challenge and to offer a theory of causation which successfully accounts for apparently problematic cases from the literature: cases of pre-emption, causation by omission, apparent failures of transitivity and cases of causal overdetermination. Along the way I will make assumptions and accept consequences that others would not. In the end, however, we will have something which we do not have now: a naturalist theory of causation which can handle the problem cases. Once we have such a theory we will be in a position to 'measure the price' as Lewis put it: consider the costs of accepting the theory, the benefits, and to compare it to rival theories. I will explicitly aim to measure the price in Chapter 9.

1.2 Lewis

I will take as my point of departure the highly influential theory of causation given in 1973 by David Lewis. According to this account:

> Actual event c is a cause of a distinct actual event e iff there is a chain of causal dependence which connects c to e, where e causally depends on c iff the following conditional is true: $\neg Oc \mathrel{\Box\!\!\rightarrow} \neg Oe$.

Here Oc represents the proposition that c occurs and the $\Box\!\!\rightarrow$ symbol represents the counterfactual conditional, so the conditional is to be read as: were it not the case the c occurred, it would not have been the case that e occurred. So, e causally depends on c if c occurs, e occurs, and if c had not occurred, e wouldn't have.

What does it mean to say that if c had not occurred, e would not have occurred? In other words, what are the semantics for the counterfactual conditional? Working out a viable semantics for counterfactuals was a significant philosophical breakthrough in the 20th century and it is a breakthrough for which Lewis [2001] shares some credit with Robert Stalnaker [1968]. Lewis and Stalnaker both offered a semantics based on the notion of possible worlds. Possible worlds are ways the world could have been—every way you can imagine—and these worlds are to be thought of as ordered or ranked in terms of their similarity to the actual world. The worlds that are more like ours in matters of fact and in their laws are *closer* than worlds which are less like ours.

According to Lewis $\neg Oc \ \Box\!\!\rightarrow \ \neg Oe$ is nonvacuously[1] true iff there is some world in which $\neg Oc$ and $\neg Oe$ are true, which is closer than any world in which $\neg Oc$ is true and in which $\neg Oe$ is not. In other words it is true that if c had not occurred, e would not have occurred if and only if there is some world such that (i) c and e do not occur, and (ii) there is no closer world where c occurs and e does not.

There is a simplification available if we follow Stalnaker and accept, as Lewis did not, the Limit Assumption that there is some closest $\neg Oc$ world. If we assume the Limit Assumption, we can say: $\neg Oc \ \Box\!\!\rightarrow \ \neg Oe$ is nonvacuously true iff all of the closest $\neg Oc$-worlds are $\neg Oe$-worlds. In words: if c had not occurred then e would not have occurred is true if and only if all of the closest worlds where c does not occur, e does not occur either.

I do not believe that anything in what I say is affected by accepting the Limit Assumption and so I will avail myself of this simplification throughout without meaning to commit myself one way or another regarding the truth of that assumption.

Lewis's analysis of causation has had enormous impact on the philosophical literature and its success no doubt stems from its relative simplicity and its ability to match common sense in a wide range of ordinary cases. The purpose of this analysis was to give a broad and non-discriminatory account of causation, that is: to give an account of causation which was not relativised to selection effects or interests. It was also a reductive account, seeking to analyse causation in terms of a putatively more basic notion: that of counterfactual dependence.

However problem cases soon emerged, in particular cases where we would ordinarily assert that there was causation, yet in which there was no counterfactual dependence between the putative cause and effect. To use a well-worn example:

> **EP:** Billy and Suzy are out to vandalise. Suzy reaches for the only rock and throws it at the window. Had she not thrown it Billy would have, and he is notoriously accurate. The rock strikes the window and the window breaks.

In this case it is clear that Suzy caused the window to break and that Billy did not. So, by the counterfactual analysis it should be the case that had Suzy not thrown the rock ($\neg Oc$) then the window would not have broken ($\neg Oe$). However, because Billy would have thrown it instead, this is not true: even if Suzy had not thrown the rock ($\neg Oc$), the window would have still broken (Oe). So it seems at first glance as though the counterfactual analysis gets the case wrong. Such cases are said to have an Early Pre-emption structure, that is a structure whereby the actual cause pre-empts the back-up by cutting it off at an early point. This is important because it means that Lewis's theory can appeal to some mid-point along the path the rock took between Suzy and the window. An event that takes place at such a mid-point (call it event d) takes place after the back-up has been frustrated so, by the time d occurs there is no longer any back-up event to guarantee the window breaking. So, the window breaking

[1] Following Lewis, this is a useful shorthand for saying that there are at least some antecedent worlds ($\neg Oc$-worlds). If there were no $\neg Oc$-worlds then the antecedent would be false and the conditional would be vacuously true.

event e counterfactually depends on event d.[2] Of course, Billy's throw could not have brought about d—it was a different rock thrown at a different time along a presumably different trajectory—so d counterfactually depended upon Suzy's throwing of the rock (c). Thus there is a chain of counterfactual dependence which runs from e to d and from d to c so, by Lewis's *transitive* account of causation, c is a cause of e despite the lack of *direct* counterfactual dependence.

Such cases depend upon the transitive nature of Lewis's analysis and I will discuss this important issue in Chapter 7, however there are cases with a similar structure that do not seem open to the same response because they lack a mid-point such as d, a midpoint which is *partisan* with respect to being dependent on the cause (Suzy) and not the back-up (Billy). Such cases are known in the literature as cases of late pre-emption:

> **LP:** Billy and Suzy are out to vandalise. Each throws their own rock accurately at a window but Suzy throws faster and her rock reaches the window first. The window breaks and Billy's rock sails through the void.

Since there is no point at which Suzy's rock is not backed-up by Billy's, there is no partisan midpoint via which we can trace a step-wise causal dependence from the effect to the cause. These cases are taken to be amongst the most difficult for any causal theory to account for. In Chapter 2 of this thesis, I will lay the groundwork for the solution I offer to these cases in Chapter 3. On the basis of that solution I will develop the ACCT Analysis of causation through Chapters 4 and 5. I will then apply that analysis to three other key problem cases for counterfactual analyses of causation: what should we say about cases where it seems like absences (which are presumably not events) cause things? (Chapter 6); what should we say about apparent counterexamples to the transitivity of causation? (Chapter 7); what should we say about cases where an effect was overdetermined and so the effect depended on *neither* apparent cause? (Chapter 8). In Chapter 9 I will consider how this analysis fares against rival contrastivist and causal modelling accounts—in short, I will *compare the prices*.

1.3 Assumptions, Caveats and Terminology

In laying out the case for my analysis I will need to make certain assumptions. Some of these are innocent and harmless and some require suspending concerns about a controversial matter until some future discussion. Here I make as many of these assumptions clear as I can.

Firstly, I will be assuming rather than arguing that a Lewisian counterfactual analysis is prima facie plausible. I do not mean that it is plausible *given* the problem cases such as late pre-emption, but rather that it is plausible enough for the problem cases

[2]This step requires that we read the counterfactual in what is known as a *non-backtrackting* way. This means that we hold the past relative to the event in question fixed when we consider what would have happened if it didn't occur. For discussion of this controversial reading see Lewis [1979].

to stand out within an otherwise successful theory. That will be important as I will be starting from Lewis's theory and amending it in response to problems as I go through.

I will also be assuming determinism for the purposes of the discussion. This is certainly controversial within metaphysics in general, but it is fairly standard in the counterfactual causation literature as it makes the examples easier to sketch and the semantics easier to state. Should the analysis on offer succeed within a deterministic framework then that should motivate attempting to extend the theory to the indeterministic cases too so this is an assumption of convenience, not a central commitment of the theory.

I will not be taking a stand on modal realism. It would not trouble me to be committed to the *reality* of the possible worlds that are discussed here—I do not find the idea as unpalatable as some—but I do not believe that I need to share Lewis's (in)famous commitment [1986c] to modal realism.

Regarding the issue of which worlds are close and which are not, I think it important to note that we may well agree on a particular similarity criterion for possible worlds yet still disagree on which of those was the closest $\neg Oc$ world, precisely because we disagree about the status of c and what it would be for c not to occur. I will discuss this sort of issue at length in Chapter 2 and again in Chapter 4 and apply my findings throughout the thesis. However, whilst I do not deny that our standards of similarity for worlds will vary by context, much as Lewis says in his [1979], I will proceed as though we can take the ordering to be fixed by some objective standard once the context is fixed. This is important for establishing a mind-independent standard of causal connection as my aim (II) requires. Lewis's use of a closest worlds metric is highly controversial, and adding this objectivity requirement may make it even more so, but if causation is a mind-independent, objective feature of our world that reduces to counterfactual dependence between distinct events, then the truth conditions for those counterfactual dependence relations had better be mind-independent and objective too.

I will be assuming four-dimensionalism, the view that time is simply another dimension which is perfectly analogous with the traditional three dimensions of space. Just as physical locations such as the Himalayas are distant to me spatially (i.e. distant in the three dimensions of space), past and future versions of the Himalayas are distant to me spatio-temporally (i.e. distant in the three dimensions of space *and* the fourth dimension of time). Thus regions of spacetime in a world pick out a four-dimensional region in that world. For a full exposition and a compelling case for four-dimensionalism see Sider [2001]. The final view I will advocate does not strictly require this assumption but my argument to that conclusion is easier to make if objects and events are close analogues, as I take it they are under four-dimensionalism.

I will assume that events are the causal relata. This is not uncontroversial[3] but it is common in the counterfactual causation literature. Less common is the further assumption that events are world-bound individuals, identical with the region of spacetime that they occupy. The merits of this approach to events will be laid out in Chapter 2.

[3]For alternative accounts see, for example, Mellor [1995] and Woodward [2003], and for a general discussion see Ehring [2009].

I will be assuming that there is a semantic/pragmatic distinction that roughly tracks the difference between the truth conditions and assertability conditions for a sentence. This will be important throughout but especially in Chapters 3 and 6 where I make use of the difference between giving the truth conditions for a causal claim and giving assertability conditions.

It will be useful for me to refer to *causal claims*, *causal facts* and *causal expressions* throughout the discussion. I take causal expressions to be ordinary spoken (or written) assertions of the form 'c caused e'. I take causal claims to be somewhat distinct: they are propositions concerning the relations of events in the world. The causal facts are just those propositions which are true. Those views of causation which take the causal facts to be determined by mind-independent and objective factors, I will refer to as *realist* or *naturalist* invoking Strawson's conception of 'natural'.

1.4 Method

It is worth remarking on the method to be adopted here at the outset. Standardly, theories of causation are offered and then tested against their ability to conform with intuition on a range of imagined examples. I will be attempting to offer an account of causation which at times requires that I reject certain causal intuitions (that others have claimed) or require that certain claims which some find counter-intuitive ought to be accepted. So, when is intuition a guide and when is it not? Lewis offered the following:

> If one event is a redundant cause of another, then is it a cause *simpliciter*? Sometimes yes, it seems; sometimes no; and sometimes it is not clear one way or the other. When common sense delivers a firm and uncontroversial answer about a not-too- far-fetched case, theory had better agree. If an analysis of causation does not deliver the common-sense answer, that is bad trouble. [Lewis, 1986e, p.194]

The problem with this outlook is that it defers all-too-thoroughly to intuition and leaves no room for a prescriptive theory. Here Lewis sounds like he is endorsing a sort of Canberra Plan approach to causation where if we simply list the causal platitudes endorsed by the folk, we will have derived a (likely disjunctive) theory of causation. This is an approach which Lewis would later explicitly reject [2004a, p.76], but the earlier statement quoted above indicates that philosophers of causation should defer to intuition almost without challenge. Hall responds to this idea exceptionally well in my opinion and I will quote his response at length here as it serves as a powerful statement of the methodological sensibilities that I will be guided by in this thesis:

> Why *not* accord intuitions about cases such a high degree of respect? Because a sensible metaphysical position is that facts about what causes what *reduce* to facts about the complete history of physical states the world occupies, together with facts about the fundamental laws that govern the evolution of these states.... Accept this reductionist picture—as I do, and

> as most authors in the counterfactual tradition seem to, either implicitly or explicitly—and it seems that even perfect success at "triangulating" on intuitions about cases will accomplish nothing more than the production of a semantics for a fragment of English. Why should scientists, philosophers of science, or metaphysicians care about that?
>
> They shouldn't. But it doesn't follow that they should not care about intuitions *at all*. That would be an overreaction. Rather, they should treat intuitions about cases as defeasible evidence of the existence of a theoretically useful concept, worth careful articulation and study. This is, I think, a sensible attitude to take towards many topics in science and philosophy. Do our firmly held intuitive judgments involving the word "knowledge" track any concept of genuine interest for epistemology? Do our firmly held intuitive judgements involving the word "life" track any concept of genuine interest to biology? And so on. It's quite difficult to answer these questions well. But it seems clear that the best way to approach them is to *start out* with the assumption that trying to produce an account that respects the given intuitions will lead to something worthwhile.. . .
>
> The shift from viewing intuitions as non-negotiable data to viewing them as "guides" makes a difference to the dialectical role of examples. It won't do to exhibit some example, point out that [certain theories] get it wrong, and declare them refuted. Rather, rejecting them on such a basis only makes sense if one can produce a *better* account, and say *why* it is better, beyond its ability to more closely fit the intuitive data. [Hall, 2007a, p.2-3][4]

Following Hall I will be taking intuitions to be evidence, but defeasible evidence, in favour or against a theory of causation. A theory which deviates from a widely held intuition bears the explanatory burden to say why, and a theory which conforms with intuitions on a given case better than another theory does is not automatically a better theory overall. In such cases we must do as Lewis says and 'measure the price'. The cheapest theory wins.

[4]This quote is from a pre-print of Hall's eventual [2007b] paper published in *Philosophical Studies*. The methodological sermon appears in the pre-print that I cite here but not in the final printed edition but Hall considers the pre-print the official version (personal correspondence). The print edition was shortened due to space constraints.

2

Event Modality and Causal Contextualism

> How much delay or change do we think it takes to replace an event by an altogether different event, and not just by a different version of the same event? An urgent question, if we want to analyze causation in terms of the dependence of whether one event occurs on whether another event occurs. Yet once we attend to the question, we surely see that it has no determinate answer. We have not made up our minds; and if we presuppose sometimes one answer and sometimes another, we are entirely within our linguistic rights. [Lewis, 2004a, p.186]

In this chapter I argue that the contextual sensitivity attributed to certain causal claims can be traced to shifts in how we represent the events involved in those claims. I will adopt a counterpart-theoretic view of events and show that this provides a compelling bridge across the fine-grain/coarse-grain dichotomy found in the literature on event ontology. Next, using a recent proposal from Schaffer [2012a] as a counterpoint, I will argue that combining this counterpart theory and a simple counterfactual analysis of causation provides a neat fit and a parsimonious semantic treatment of context sensitive causal claims. This relates to aim (I) of my thesis: to account for our everyday causal assertions.

2.1 Event Variation Across Contexts

If events are the relata of causation, as is widely assumed, then those offering accounts of causation had better specify what they take events to be. Counterfactual accounts of causation in particular owe an account of what it is for an event to occur, and importantly what it is for a given event *not* to occur.

In this section I propose that events should be viewed counterpart-theoretically. I will then argue that our *de re* modal attributions concerning events can vary with

context. This idea is familiar from Lewis's discussion of objects and what he referred to as our *inconstant* representation of them. In §2.2 I will show how this idea of inconstancy can help us understand the context variation we find in our causal talk.

2.1.1 Event Modality

The literature on the ontology of events is split on the topic of *grain*. On the one hand you have a *fine*-grained conception of events offered by, for example, Kim [1973] (and Lewis [1986d, p241-269]) and on the other you have a *coarse*-grained conception of events offered by, for example, Davidson [1963, 1969].

On the Kimian view events are constituted by the triple [object, property, time]: an object having a property at a time. Thus, an object at a given time can constitute one event in virtue of one property and constitute a second in the same place, at the same time, in virtue of some *other* property. The non-constitutive properties that are present in the object at that time are said to be 'exemplified' by the event but not constitutive of it. Such events are considered fine-grained because there are as many of them in a given space-time region as there are properties, leading to a fine discrimination between events in a world.

On the Davidsonian view events are individuated by their causal role and are extensional. This means that they can be re-described, via different predicates, *salva veritate*. According to this view the event *Bill's birthday party* is identical with *the party in the penthouse*. Since there was only one party, the idea that there is just one event, twice represented, has *prima facie* plausibility. Davidson's view of events offers a coarse discrimination between events since it posits far fewer events[1] in any given space-time region than the Kimian alternative.

So, a key difference between the fine-grainer and the coarse-grainer lies in how many events they posit in the actual world: the fine-grainer posits vastly more than the coarse-grainer. These differing individuation conditions are clear enough in the actual world, but our counterfactual considerations add a modal dimension to events—had Bill's party been held on the ground floor instead of the penthouse, would it have been the same party? Nothing in the individuation conditions of events which I have introduced commits either theory to a particular answer to this *counterfactual* question.

The counterfactual question requires us to consider two occurrences—one in the actual world and one in another, possible, world—and ask whether these two occurrences constitute the *same* event. For Kripke [1980], individuals can exist in many worlds at once and so a Kripkean about event modality might think that it is possible for these two occurrences in different worlds to constitute the same event. For Lewis, individuals were world-bound and so could not *literally* exist in multiple worlds, but they could have *counterparts* in those worlds instead (more on this below). Oddly, Lewis took events to be different from objects or people in this respect and he argued for a transworld identity of events.[2] In this chapter in particular, and in the thesis

[1] "Perhaps just one: I am uncertain both in the case of substances and in the case of events whether or not sameness of time and place is enough to ensure identity" [Davidson, 1969, p.306].

[2] One might reasonably expect that a four-dimensionalist such as Lewis, who thought that objects, defined by their extension in spacetime, are individuals, would also think that events which are

more generally, I will, unlike Lewis, take events to be *individuals* but, like Lewis, I will take individuals to be world-bound particulars. That means that I take Bill's party in this world to be one event and any occurrence in another world, no matter how similar to Bill's party, to be a different event.

Given this understanding of events, the question about whether or not the party would have been the *same* if it had occurred in the basement cannot be a question of literal identity on pain of triviality. The question is not trivial, so some weaker form of sameness must be applicable. The sameness under consideration is the sort that we find in our standard modal ascriptions: I could have been taller, the pylon could have toppled or the fire could have been contained. In each such case we are saying that there is a possible scenario in which I, the pylon or the fire exist, but exist in some altered state—taller, toppled or contained respectively. Each entity is considered *the same* despite the alteration. The counterfactual question about Bill's party asks whether it would have been the same party under a specified alteration, namely having taken place on the ground floor instead of the penthouse. To settle this question we require not only actual-world individuation conditions but cross-world, or *modal* individuation conditions too. Let us follow tradition and call those features that an object or event must have to be considered the same across worlds as forming the *essence* of that object or event. The features that are not essential, we will call *accidental*. The crux of our counterfactual question then is this: is the location of Bill's party essential or accidental?

I said that neither the fine-grainer nor the coarse-grainer (as characterised) were committed to a particular view of the modal individuation of events, but one may think that we can infer an answer on behalf of the fine-grainer who already splits their events into those features which are constitutive and those which are exemplified. Mapping this dichotomy onto the modal notions of essence and accident seems natural—if an event is individuated by the presence of its constitutive properties in the actual world, it would make sense that these same properties individuate the event modally. Thus the fine-grainer has a conditionalised answer to the counterfactual question: if the penthouse location of Bill's party is constitutive of the event, then it is no longer the same event on the ground floor, however if the location is merely exemplified, then the possible event that takes place on the ground floor can be considered the same event. This conditionalised answer now prompts the question of which features of events are constitutive and which are merely exemplified?[3] This question will get a different answer for every one of the countless events that occupy a given region and without knowing which of those events is under consideration, we will be no closer to having

similarly extended in spacetime, are also individuals. In his clearest articulation of his event ontology, his paper *Events* [1986d, p.241-269], Lewis argued for transworld, rather than counterpart-theoretic events Why he thought this is not obvious from what he says in that paper so I think this is an interesting question for Lewis scholars, though beyond my scope here.

[3]Kim [1973] argues that the subsumption of events under laws helps establish which features are constitutive. I will ignore this response here for two reasons: First, I wish to present a more neutral fine-grained account without this further commitment. Second, it is far from obvious how Kim's view could fit with the counterfactual account of causation that I will be considering later.

an answer to the counterfactual question.[4]

It is even less obvious what the coarse-grainer should say. The coarse-grainer gives no priority to one set of features over another in the constitution of the event and so there is no natural mapping of such features onto the essential/accidental split. Three options present themselves: (i) treat all of the properties as essential, (ii) treat none of the properties as essential, (iii) treat some of the properties as essential (and the others accidental). Option (i) denies our standard modal attributions about the pylon, the fire and me—all are strictly false. Option (ii) accepts our standard modal attributions about the pylon, fire and me but problematically accepts *any* modal attribution whatsoever. On this view 'that pylon could have been an electron' is true. This just makes a nonsense of possibility attributions in general. Option (iii) allows the acceptance of some modal attributions and the rejection of others—as we might have hoped—but absent a principled way of telling which features are essential or accidental, we still cannot tell the true attributions from the false. We are no further forward in answering the counterfactual question.

So, on the one hand we have a fine-grained account of events which has a natural modal reading but a bloated ontology and on the other hand we have the coarse-grained view which has a more parsimonious ontology but no natural way of reading our modal claims. The problem of answering the counterfactual question about Bill's party cuts across both accounts.

2.1.2 Event Counterparts

There is strong precedent in the realm of objects for handling our counterfactual question. Where a strict standard of sameness (identity) restricts claims of sameness to only those objects who share all and only the same features, a weaker standard is employed when we entertain certain counterfactual claims—we will happily say *it could have been larger* or *it could have been green*. One strategy which allows us to accommodate both standards is to adopt a *counterpart theory* for objects.[5]

The object before me is a keyboard. There is only one keyboard which is *exactly* the same as this keyboard: this very one. No other keyboard, no matter how qualitatively similar to this one, is identical with it. However that does not preclude the keyboard from having counterparts in other possible worlds where things are different. In worlds without plastic perhaps there is a functionally similar keyboard made of some natural resin. In such a world the counterpart keyboard does not strictly speaking sit on this desk in front of me, but rather, it sits on a counterpart desk in front of a counterpart of me in that other world. Objects such as the keyboard, desk or me, stand in a counterpart relation with objects in other worlds such that these counterparts are sufficiently similar to be counted as representing the object in that world, without strictly being the object in that world. In counterpart theory we take the truth of *de re* modal attributions to depend on both the object and its counterparts. Possibility attributions are true if they are true of some counterpart and false otherwise.[6] For

[4] I will discuss what clues the expression of the event provides in §2.2.2.

[5] For a well developed version of counterpart theory see Lewis [1968a, 1986c, 2001].

[6] Note that the actual object is taken to be one of its own counterparts [Lewis, 1968a, p. 114 P6].

example, the actual keyboard is made of plastic but *this keyboard could have been made of a natural resin* will be true if and only if there is a counterpart of the keyboard in some world w that is made of natural resin. On the other hand necessity attributions ('must', 'essentially', 'necessarily') applied to an object are true only if the attributed feature is to be found in all of its counterparts, and false otherwise. For example, *the keyboard is essentially white* will be true if and only if all of the counterparts of the keyboard are white.

The advance offered by counterpart theory is that it resolves the apparently inconsistent claims about identity: strict identity is maintained, but less strict modes of identity are tolerated via the counterpart mechanism.

I propose that events should be understood as *individuals* within counterpart theory. Strictly speaking the event is just the concrete, world-bound, particular that occupies a given space-time region—this is the actual world individuation that the coarse-grainer is advocating. In our talk of event modality, however, we employ counterpart relations that satisfy our modal attributions of those events. Thus, had the party occurred on the ground floor, it would have counted as the same party via being a counterpart to the actual (penthouse) party. Strictly speaking there was only one party, but there is a sense in which it would have remained the same had it occurred differently.

Embracing a counterpart theory of events offers an explanatory bridge between the coarse-grain and the fine-grain views. If events were transworld entities, literally existing in multiple different worlds, then the distinct modal attributions we could make of the party—it could have been louder, it could have been dull—would imply that distinct events (the loud party, the dull party) overlap in the actual world, just as the fine-grainer maintains. However, on a counterpart-theoretic view of events these distinct attributions do not imply distinct events that overlap, but rather distinct representations of the single individual in the actual world. In different contexts, counterpart theory can support each of the different fine-grained attributions: under one counterpart relation Bill's party has a loud counterpart (but no dull ones), under another it has a dull counterpart. The strength of counterpart theory is that it can achieve such fine-graining without bloating our ontology because there remains just one, re-describable, event in the actual world that corresponds to each of these different attributions. Thus, the proposed counterpart theory of events is a coarse-grained view which can track our fine-grained event attributions.

2.1.3 The Inconstant Modality of Events

Counterpart-theoretic events allow for the double-standard of sameness which seems to resolve the underlying conflict in event ontology. In Lewis's original formulation of counterpart theory [1968b] there was just one such counterpart relation but in subsequent iterations [1983a, 1983d, 1986c, 2003], Lewis held that there were indefinitely many counterpart relations that could hold accross different contexts. This means that adopting a flexible counterpart theory, as I do, doesn't by itself resolve which of two

competing *non-strict* counterpart relations should be taken to apply in a given case.[7]

Perhaps Bill would not think it would have been the same party without his friends being present and yet his neighbours care only that it was noisy, not who was the source of the noise. To them the party with a totally different set of members could be the same event. These event identity standards are in conflict—the event cannot at once be the same and not the same under a given alteration—so which is it to be?

I think this question is misleading. Whilst Bill and his neighbours are both talking about the same party—the one that actually took place—they are talking of it under different counterpart relations. One relation, the neighbours', contains a counterpart with completely different party-goers and the other, Bill's, does not. The different counterpart relations make for a difference in the acceptability of certain *de re* modal attributions concerning the event. Nonetheless, they are talking of the same actual-world party.

Similarly for the delayed concert: the tickets remain valid and the band plays the same songs in the same order, so for many it will be the same event. However the back-up band may be different, or the lighting engineer or perhaps it is re-scheduled for a night when a certain fan cannot go. To those who experience it, or fail to experience it, in a way that is peculiar to the re-scheduled event—that is, they experience it as essentially having some feature that would have been absent from the originally scheduled concert—this is not the same event at all. For one group of people it is the same event, and for another group it is not, so there are different standards of sameness in different contexts. There are a multitude of counterpart relations that could apply to any given event and context shifts which applies and when.

A final example: A train can travel down *Local*, *Express* or *Broken* lines to the station arriving, respectively, on time, early or not at all. When the train travels down Local, is that event essentially the taking of the Local line? Or is that event essentially the taking of *some* functional line, but only accidentally the taking of the Local line? The signaller may well adopt the former view as it corresponds with the precision of his intentions and has a knock-on effect concerning traffic on the various lines. However a nervous passenger may only care that the line is functional, especially if that passenger is unaware of the Express line, but all too well aware of the Broken one. This passenger will treat the event as having a different essence. Which essence is the right one? As Lewis says in the opening quote: *once we attend to the question, we surely see that it has no determinate answer.* Both the signaller and the passenger are within their linguistic rights to assert competing *de re* modal attributions regarding the train's journey.

To recap, my proposal is to apply a counterpart theory, typically applied to objects, to events. In object-involving events, this a natural move as the *de re* modality of the objects within the event would seem to impact the *de re* modality of the event as a whole. The examples given above, however, (the party, the concert and the train) appear to vary with context quite apart from the modality of the objects that they involve: it is not Bill's counterpart relation that varies, or the performer's or the train's,

[7]This is the same issue as the fine-grainer faced when asked about which features constituted the event — they have as many different answers as there are events so how do they choose between them?

but rather something over and above that object variation. This indicates that event expressions, and not just the objects they involve, trigger *inconstant* (to use Lewis's phrase) modal attributions and counterpart theory allows us to express this inconstancy in terms of shifting counterpart relations.[8]

2.2 Event Modality and Causal Claims

In the previous section I argued that event expressions are sensitive to contextual variations insofar as the context affects which *de re* modal attributions we will accept of that event, which is just to say that context can shift the counterpart relation that we take the event to fall under. In this section I will demonstrate the impact of shifts in an event's modality upon the truth of causal claims involving that event. I will introduce a simple counterfactual test for causation, then I will show that shifts in the modality invoked by an event expression shift the truth value of counterfactual conditionals involving these events. I will then argue that this shift accounts for certain sensitivities (contextual, sentential) in classic examples of causal contextualism.

2.2.1 Counterparts and Counterfactual Sensitivity

I will adopt the following simple causal test:[9]

> For any distinct actual events c and e, c is a cause of e if and only if c and e are linked by a chain of counterfactual dependence where e counterfactually depends upon c iff:
>
> $\neg Oc \mathrel{\Box\!\!\rightarrow} \neg Oe$
>
> In words: if c had not occurred, e would not have occurred.[10]

The truth conditions for this counterfactual conditional operator are given in Lewis [1973, p.560-561]: $\neg Oc \mathrel{\Box\!\!\rightarrow} \neg Oe$ is nonvacuously true *iff* some world where c does not occur and where e does not occur is closer than any world where c does not occur and e does. We can harmlessly simplify this for the purposes of my discussion: $\neg Oc \mathrel{\Box\!\!\rightarrow} \neg Oe$ is nonvacuously true *iff* all of the closest $\neg Oc$-worlds are $\neg Oe$-worlds.[11]

[8]The version of counterpart theory that I adopt here is left open to more than one interpretation. This is so that certain contentious commitments of specific views (Lewis's modal realism being a prime example) do not complicate the issue unnecessarily.

[9]The simple test I offer is sufficient for the cases I will discuss, but it is highly vulnerable to counterexamples in a way that more sophisticated versions may not be. See Lewis [2004a] for a run-down of the issues that plague such an account. I offer it here as an indicative test, not as a general analysis of causation.

[10]For reasons that will become plain in Chapter 4 I will **not** follow convention and drop the O. Lower case italicised letters are to signify events, except w which will signify a world.

[11]This simplification implies that there is such a thing as *the* closest possible world. Lewis thought this 'Limit Assumption' was unwarranted and so he gave the wordier locution. I will use the neater phrasing without meaning to commit either way to whether or not the assumption should adopted.

Possible worlds are possible ways the world could have been—all the possible ways—and such worlds are ranked for closeness by a similarity metric. So, the closer possible worlds are those worlds more like ours (the actual world) in matters of fact and natural law.[12]

On this account what makes c a cause of e is that e *counterfactually depends* upon c—that is, without the cause the effect would not have occurred—or that there exists a chain of such dependence between c to e. This is by no means the end of the story[13] but this basic test lies at the heart of a range of contemporary counterfactual-based accounts of causation. Such simple counterfactual analyses can have few serious defenders remaining given the apparent failures to analyse problem cases such as pre-emption and prevention, which may make it seem like an odd choice of test to employ. However whilst few would defend the idea that such a test is decisive, it remains widely accepted that it is, at the very least, strongly *indicative* of causal connectedness in standard cases. I will consider what impact the context sensitivity of event expressions has within this causal test before linking these findings to a recent proposal by Jonathan Schaffer in the next section.

Here is the main claim of this section: varying what counts as a counterpart of an event can vary which counterfactual conditionals concerning that event are true.

Let us begin with shifts in the counterpart relation. Suppose that McEnroe has just served but that the serve was awkward. Perhaps the event expression 'McEnroe's serve' invokes two different counterpart relations in two different contexts: in the context of an inattentive observer who didn't see the serve, the counterpart relation includes counterparts of the serve which are awkward and counterparts which are graceful, but in the context of the attentive coach, the counterpart relation includes only those counterparts which are awkward. By the first relation the same serve can be taken to occur in worlds where it is graceful, but by the second it cannot. Thus, shifting the counterpart relation invoked shifts which worlds that event is taken to occur in.

Recall that the truth conditions for the counterfactual conditional $\neg Oc \ \Box\!\!\rightarrow \neg Oe$ state that for $\neg Oc \ \Box\!\!\rightarrow \neg Oe$ to be true, it must be the case that all of the closest worlds where c does not occur are worlds where e does not occur. So, shifting the worlds in which the cause event is taken to occur, will affect the truth of counterfactual conditionals involving that event, when that shift alters whether or not all of the closest possible $\neg Oc$-worlds are $\neg Oe$-worlds. So, if the counterpart relation shift alters the set of closest $\neg Oc$-worlds such that either (i) they are no longer all $\neg Oe$-worlds, or (ii) they are now, but were not before, all $\neg Oe$-worlds, then that shift in counterpart relation changes the truth of the conditional from true to false in (i), and from false to true in (ii).

Relatedly, shifting the worlds in which the effect event is taken to occur, will affect the truth of counterfactual condtitionals when that shift alters whether or not all of the closest possible $\neg Oc$-worlds are $\neg Oe$-worlds. Assume that world w is one of the closest $\neg Oc$-worlds. There are two types of counterpart relation that could apply: (iii) relation

[12]For more detail on how matters of fact and law are to be weighed, see Lewis [1979].

[13]For a comprehensive discussion of a range of issues that arise for counterfactual accounts of causation view see John Collins [2004].

r_1 on which e does not occur in all of the closest $\neg Oc$-worlds, including world w, and (iv) relation r_2 on which e does not occur in all of the closest $\neg Oc$-worlds *except* world w. In the language of counterfactuals, w is an $\neg Oe$-world by (iii), but is an Oe-world by (iv). Shifting the counterpart relation from r_1 or r_2 has shifted the counterfactual conditional $\neg Oc \mathrel{\Box\!\!\rightarrow} \neg Oe$ from true to false on Lewis's semantics.

If this is correct, then shifting the counterpart relation can shift which conditionals involving events that fall under that relation are true. The counterfactual test for causation that I am employing takes c to be a cause of e if and only if a certain counterfactual conditional ($\neg Oc \mathrel{\Box\!\!\rightarrow} \neg Oe$) is met. Since shifting the counterpart relation of an event can shift the truth value of counterfactual conditionals, then shifting the counterpart relation of an event could shift the causal status of that event on a counterfactual test of causation.

In the next section I will show that certain features of our causal talk can be understood as stemming from counterpart shifts in the events involved.

2.2.2 The Sensitivity of Causal Claims

I will consider three classic cases in which shifts in the representation of an event shift the causal intuitions. In each case I will argue that there is an implicit shift in the counterpart relation being evoked across the different representations.

The first is from Hitchcock [1996]. Consider the following event:

1. Susan's stealing the bicycle.

Now consider how the causal implications change as we introduce emphasis on different parts of the event phrase:

2. Susan's *stealing* the bicycle caused her to be arrested.

This appears to be true, whereas the following appears to be false:

3. Susan's stealing the *bicycle* caused her to be arrested.

It seems that emphasising 'stealing' in (2) and emphasising 'bicycle' in (3) each shift the acceptability of the causal claim that Susan's stealing the bicycle caused her to be arrested. What is the emphasis shifting?[14] My proposal is that shifts in the event expressions indicate shifts in the counterpart relation being invoked.

If we take the emphasis to indicate the essence of the event, then we can see that in (2), all of the counterparts to the cause, c, will involve stealing—stealing of a bicycle or skis or whatever else. The closest $\neg Oc$-worlds will be those where no such c counterpart, i.e. where no such stealing event, takes place. All else being equal, the closest $\neg Oc$-worlds will not feature Susan's arrest, e, and so those closest $\neg Oc$-worlds will be $\neg Oe$-worlds too. By the counterfactual test, this means that c is a cause of e and that (2) is true.

[14]Hitchcock argues that the emphasis implies an alternative, contrast, event. I am proposing a different solution and I will say more on contrastive approaches when I come to discuss Schaffer's proposal in the following sections.

Turning to (3), it seems that all of the counterparts to the cause, c, will involve a bicycle: stealing of a bicycle, riding of a bicycle etc. The closest $\neg Oc$-worlds will be those where no such c event, i.e. no such bicycle-involving event, takes place. All else being equal, the closest $\neg Oc$-worlds will still bring about Susan's arrest, e, because the closest such worlds contain counterparts of c which are still stealings, just not of a bike (perhaps of skis). So some of those closest $\neg Oc$-worlds will be Oe-worlds. By the counterfactual test, this means that c is *not* a cause of e and that (3) is false.

So, counterpart theory, coupled with a counterfactual test for causation (call this coupling **CCT** for Counterpart-theoretic Counterfactual Theory of causation), can match intuition on this emphasis-shift case.

The next case uses the earlier example of McEnroe's serve and is taken from McDermott [1995]. This case concerns the impact of altering the description that is used to pick out the event and illustrates the impact of re-description of the event: different counterfactual conditionals hold. This first sentence seems acceptable:

4. McEnroe's tension caused him to serve awkwardly.

Yet, when we remove the adverb, we change the acceptability of the claim:

5. McEnroe's tension caused him to serve.

The difference between (4) and (5) is the removal of the adverb 'awkwardly'. Presuming that there only was one serve being discussed, the change in description may appear innocuous. Presumably to 'serve' could be to do so gracefully, but to 'serve awkwardly' could not, so there are counterparts of the serve in (5) which are awkward, and ones which are graceful, whereas there are only awkward counterparts to the serve in (4). The re-description is not so innocuous after all.

If we take the effect-side event to be essentially awkward, as I take it the first description implies, then there will be some $\neg Oe$-worlds—the closest ones—where McEnroe still serves, just not awkwardly. The closest worlds in which McEnroe is not tense are worlds in which he is still primed to serve, just absent the tension. So, we can expect that all of the closest $\neg Oc$-worlds (not-tense worlds), will turn out to be $\neg Oe$-worlds (not-awkward worlds). That makes the first claim true on a counterfactual test.

If we take the effect-side event to be essentially a serve, as the second description implies, then all of the $\neg Oe$-worlds will be worlds without a service. The closest not-tense ($\neg Oc$) worlds remain exactly like the actual world in other respects—McEnroe throws the ball and arches his back just as he does in the actual world. At least some of the closest such $\neg Oc$-worlds will yield a serve and so will be Oe-worlds too. Thus, the second claim fails the counterfactual test for causation.

Once again, armed with a sensitivity to the counterpart shifts implied by the different descriptions, the CCT account tracks our intuition. The claims that the counterfactual test says are true, we intuit as acceptable, those which the counterfactual test considers false, we intuit as unacceptable. I take these examples to illustrate the impact that implicit counterpart shifts can have on our causal attributions, at least on a counterfactual view of causation.

I take my third case from Achinstein [1975]:

6. Socrates's DRINKING HEMLOCK at dusk caused his death.

7. Socrates's drinking hemlock AT DUSK caused his death.

The first is acceptable but the second is not. Once again, there is but one drinking of hemlock but when the focus shifts to AT DUSK then *that* becomes the essential property of the event. A shift in the essential property of an event *just is* a shift in the counterpart relation it invokes and, as we have seen, shifting the counterpart relation indicates a shift in the set of acceptable *de re* modal attributions of the event.

If we take the cause-side event to be *essentially* a drinking of hemlock, as I take it the first focus implies, then the closest $\neg Oc$-worlds will be worlds where Socrates does not drink hemlock. Absent the drinking of hemlock Socrates will not die,[15] so all of these $\neg Oc$-worlds will be $\neg Oe$-worlds. This means that the first claim passes the Lewisian test for causation.

If we take the cause-side event to *essentially* take place at dusk, as I take it the second focus implies, then the $\neg Oc$-worlds are those in which a counterpart poisoning does *not* occur at dusk. All else being equal, any world where Socrates drinks hemlock will be closer to the actual world than a world where he does not. So, the closest $\neg Oc$-worlds will include Socrates drinking hemlock—just not at dusk—and subsequently dying. So, at least some of the closest $\neg Oc$-worlds will be Oe-worlds and so the second claim fails the Lewisian test for causation.

This shift in emphasis is just like that of the Hitchcock example above concerning Susan's theft: in virtue of shifting the counterpart relation of the event in question, the emphasis shifts the truth value of the causal claim on the Lewisian account. Thus, the Lewisian semantics along with counterpart theoretic events—the view I dub CCT—tracks the intuitive acceptability and non-acceptability of 6 and 7 respectively.

So, for clarity, here is the CCT view: c is a cause of e iff (i) c and e are distinct actual events; and (ii) c and e are linked by a chain of causal dependence, where causal dependence is defined as follows: e causally depends on c relative to counterpart relation x iff the following counterfactual conditional is true:

$$\neg Oc_x \boxdot\!\!\rightarrow \neg Oe_x$$

In ordinary causal discourse, the value x is set to a specific value, n, which is determined in part by the context of utterance and the mode of representation of c and e.

2.3 Causal Contextualism

In a recent paper, Jonathan Schaffer [2012a] argues that our causal claims are context sensitive and that, since no existing pragmatic mechanism exists to account for it, that

[15] Well, of course he will die eventually. The idea here is that he won't die in a *relevantly similar* way. Establishing what counts as *relevantly similar* is exactly what I take the counterpart relation to do. Just as there will be different standards of relevant similarity, there will be different counterpart relations.

variation must be semantic in nature. He goes on to propose a contrastive semantic framework that would account for the contextual variation.

In this section I will introduce Schaffer's argument for Causal Contextualism and then briefly summarise his contrastivist position. I will then argue in §2.4 that my counterpart theoretic approach offers a satisfying alternative to Shaffer's treatment of the contextual variation of causal claims.

2.3.1 Schaffer's Dichotomy

The aim of Schaffer's argument is to establish a thesis he calls *Causal Contextualism*. He defines this as follows:

> A single causal claim can bear different truth values relative to different contexts, where this difference is traceable to the occurrence of 'causes,' and concerns a distinctively causal factor. [Schaffer, 2012a, p.37].

The requirement that the contextual variation be traced to the word 'causes' means that *Causal Contextualism* will be false if the contextual variation can be traced elsewhere—say, to the events (and their counterparts). The alternative Schaffer offers is *Causal Invariantism*:

> It is not the case that a single causal claim can bear different truth values relative to different contexts, where this difference is traceable to the occurrence of 'causes,' and concerns a distinctively causal factor. Causal claims are context sensitive in their acceptability, but the context sensitivity of causal claims is a wholly pragmatic phenomenon [Schaffer, 2012a, p.40]

This is the denial of *Causal Contextualism* plus a positive requirement that the context variation be a *wholly pragmatic phenomenon*. There is a genuine dichotomy at play here: is the sensitivity to context a wholly pragmatic or is it a partly semantic phenomenon? However, the two theses presented do not track this dichotomy because there remains logical space to deny *Causal Contextualism* but reject that context variation is a wholly pragmatic phenomenon either. In particular, Schaffer overlooks semantic accounts which do not treat the context variation of causal claims as distinctively causal.

The two options presented by Schaffer are clearly not logically exhaustive,[16] but he takes arguments against *Causal Invariantism* to provide evidence in favour of *Causal Contextualism* in his discussion. This is tantamount to a false dichotomy since there is a third way being overlooked. My view of combining a counterfactual causal test with counterpart theory represents a version of this third way.[17]

The CCT view that I have been advocating amounts to a semantic treatment of the contextual sensitivity of certain causal claims: what shifts from one context to another is the *truth* of the claim, not just its assertibility. This is not a wholly pragmatic

[16] Schaffer concedes as much in [fn. 3, p.60].

[17] It is worth noting that Schaffer also explicitly endorses a counterpart theoretic view of events in his [2005].

treatment, and so CCT does not meet Schaffer's criterion to be *Causal Invariantist*. As I have presented it, the CCT view does not trace the contextual variation to the presence of 'causes' and so it should not be considered *Causal Contextualist* either.[18] Of course, if further examples emerge which demonstrate that the word 'causes' makes a distinctively context-sensitive semantic contribution, over and above the contribution of the events, then this would precipitate a modification of the view. Note, though, that any contextualist view which denies or overlooks the contextual contribution of the events owes an explanation as to why events display across-context variation independently of causal discourse, but yet do not contribute to the context sensitivity of the causal claims which they constitute.

So, CCT offers a third way between the contextualist and the invariantist that Schaffer characterises: it has the contextualist feature of truth value variation across contexts, but that variation concerns the entire causal claim and is traceable to the contribution of the event expressions. CCT takes the semantic contribution of 'causes' to remain static across contexts. I will now briefly introduce Schaffer's positive proposal before going on to argue in §2.4 that the *data* he considers can indeed be accounted for within the CCT view without any additional contribution from the word 'causes'. The aim is to establish that, despite being overlooked, CCT remains a viable alternative.

2.3.2 Schaffer's Contrastive Proposal

Having given the *prima facie* case for contextualism, Schaffer goes on to propose a semantic treatment of causal claims that traces their context variance to the presence of 'causes'.

Schaffer's proposal is that the verb 'cause' projects two contrast places in the causal claim, and that context dictates which contrasts are salient. These contrasts can be seen explicitly in the *rather-than* constructions of certain causal claims, where both the cause and the effect are contrasted with salient alternatives, but are often suppressed or implicit. For example: 'Susan's stealing the bicycle caused her to be arrested' should be interpreted as meaning 'Susan's stealing the bicycle *rather than borrowing it*, caused her to be arrested *rather than remain free*'. In this case the word 'causes' is taken to project each of the *rather-than* clauses which were suppressed in the original, more natural, formulation.

On Schaffer's view the binary surface grammar of 'c causes e' is semantically incomplete. To complete the causal claim we need to plug in alternatives to the cause and to the effect such that we get a four-place claim: c rather than c^* causes e rather than e^*.[19] Every causal claim is taken to have this quaternary structure at the level of

[18] Perhaps there is a sense in which it is traceable to the presence of the word 'causes': the counterfactual construction itself stems from presence of the word 'causes' and it is the truth of this construction which varies across contexts. Of course the word 'causes' and the construction that it gives rise to are present in contextually variant causal claims, but that is just because they are present in all causal claims. I am proposing that the context variance is contributed by, and hence traceable to, the events alone—at least in the examples considered. I thank Daniel Nolan for pointing this out.

[19] To mirror Schaffer I will treat the contrast as being a *specific* alternative to c (represented by c^*) rather than a set of alternatives (which are represented by C^*). Schaffer suppresses this distinction in his [2012a, p.45] for simplicity and so when discussing Schaffer's contrastivism I will follow suit.

logical form and by filling each of the places for c, c^*, e and e^* the contextually variant elements, i.e. the contrasts, are fixed. Thus, the explicitly contrastive formulation is context invariant.[20]

So, Schaffer offers a contrastive rendering of 'c causes e' and whilst he stops short of committing to a full analysis of this rendering he does gloss it as follows: c rather than c^* causes e rather than e^* iff (roughly) if c^* had occurred, e^* would have occurred [p.46]. For the purposes of this discussion, then, we can sketch Schaffer's position as follows: 'c caused e' is true iff had c^* occurred, e^* would have occurred (where c^* and e^* are supplied by context).

I will next argue that the CCT view can handle the context variation cases as well as the contrastive view can, but I will also argue that tracing the contextual variation to the inconstant modality of the events, not the presence of 'causes', is more parsimonious and has independent motivation.

2.4 Contrasting with Contrastivism

In this section I will argue that the CCT view can account for the contextual variation found in the classic cases that Schaffer discusses. I will not only argue that CCT can match the results of contrastivism in these cases, but that it can be expected to match the contrastivist's results more generally. The aim of this section is to establish that the CCT view is, at the very least, a viable alternative to contrastivist treatments of the context sensitivity of causal claims.

2.4.1 Sentential and Contextual Sensitivities

Within his [2012a] paper, Schaffer groups the bicycle, McEnroe and Socrates cases that I discussed in §2.2 as part a class of *Sentential Sensitivities*.

In discussing each of these cases, Schaffer makes the following caveat: "...unless one has an implausibly fine conception of events..." [Schaffer, 2012a, p.39]. This caveat seems to imply that the contextual variation in these cases cannot be traced to the events since the events (the stealing and the serving) are one and the same across contexts. However, the CCT view I am proposing provides just the sort of framework that makes sense of there being just one event and for event referring terms to alter the counterpart relation they pick out relative to context. I have already argued in §2.3.1 that the variation in these cases can be traced to the events in this way.

In addition to these *Sentential Sensitivities*, Schaffer presents three cases which he describes as examples of *Contextual Sensitivity*. I will argue that the CCT view on its own can offer a semantic solution for two of these cases and I will propose that we combine CCT with the standard Gricean pragmatic maxim of *relation* to resolve the third.

Beginning with Schaffer's example of **Causal Inquiry**. The example given is:

However, when I come to talk of contrastivism more generally in Chapter 9, I will switch to the more common use of contrast *sets* (C^*).

[20] This is a crude summary. For a more detailed account see Schaffer [2005, 2012a].

8. John's boldness caused him to kiss Mary.

Schaffer argues that (8) is acceptable against the backdrop of one causal inquiry: where the question is 'Why did John *kiss* Mary?, but unacceptable against the backdrop of another: 'Why did John kiss *Mary*?'. I think that this is quite possibly right. However, the backdrop in this case, the one which contains the question, constitutes the context and it is the context that fixes which counterpart relation will be invoked on the CCT view. In this case the first version of the question, the emphasis on 'kiss' invokes a counterpart relation such that the effect is essentially a kiss and the second, with the emphasis on 'Mary', invokes a counterpart relation such that the effect essentially involves Mary. In the scenario Schaffer is proposing, where kissing is the relevant feature under consideration, then the CCT view can appeal to the same contextual clues that Schaffer's contastivist can.

Turning to the example of **Multiple Alternatives**:

9. The train's taking *Local* caused it to arrive at the station.

This claim seems acceptable when the salient alternative is Broken but it seems unacceptable when the salient alternative is Express (since the train would have arrived anyway).

If there is a feature of the context that fixes the salient alternative, then that feature is part of what fixes the counterpart relation of the event. If the salient alternative is Broken then the counterpart relation invoked in that context will reflect that—perhaps its essence involves being a train route and being not-Broken. Express shares this essence and so is not sufficiently different to constitute $\neg Oc$ despite being different in many respects from the actual world event. Such is the nature of counterparts: whilst they are similar in some respect, they can be very different in another. Again, the CCT view can appeal to the same contextual clues as Schaffer's contrastivist.

Finally to the case of **Selection**: We commonly foreground certain causes and relegate others to the status of *background conditions*. This selection varies with context as in the following example.

10. The presence of oxygen caused the forest fire.

Perhaps a visiting Venusian [Putnam, 1982], astonished by the ubiquity of such a combustible substance, would find this claim acceptable, but a forest ranger would not (preferring instead to blame the stray matches). The acceptability of the causal claim varies in relation to the background considerations at play.

Schaffer considers the Gricean maxim of *relation* [1968] as a possible explanation of the context shift: it is relevant to Venusians, but not to rangers, that the oxygen was present and so it is acceptable to Venusians, but not rangers, to cite the oxygen as a cause.

Against this notion, Schaffer claims that we will ordinarily assert the negation of (10):

11. The presence of oxygen *did not* cause the forest fire.

If we would assert the negation, then (10) was not merely irrelevant, it was outright false, implying a semantic variation across contexts.

It is not at all clear how this fits in with Schaffer's contrastivism. Take the following contrastive rendering of the **Selection** example:

12. The presence rather than absence of oxygen caused there to be a forest fire rather than no fire.

This is true on Schaffer's contrastive account quite irrespective of whether the rangers utter it or the Venusians. What Schaffer requires is that when the Venusians say (10), then (12) is indeed the appropriate contrastive to form, but when the rangers utter (10) a different, false, contrastive should be formed courtesy of their different context. What is this alternative false contrastive that emerges from the rangers' context? I see none, and nor does Schaffer (p52). Further, what contrastive truth is expressed by (11) in the ranger's context? Again, I see none and nor does Schaffer:

> Lacuna: if [10] does not receive any natural interpretation then its denial should not either, which does not quite fit that data in [11]. So it would be smoother for me to say that [10] does receive some interpretation as a contrastive falsehood in the context of the forest rangers. But I do not currently have any contrastive falsehood to suggest for the role. [Schaffer, 2012a, p.61, my numbering]

Selection just doesn't fit with his overall picture.

As it stands, the CCT view cannot account for the contextual variation in this case either. The fire counterfactually depends on *both* the presence of oxygen and on (say) the misuse of the matches, and I can see no reason to think that context is shifting the event modality in such a way as to shift the nature of that dependence. In short: both are causes. CCT therefore counts (11) as false.

I suggest that the problem with **Selection** is that accepting (11) amounts to begging an intimately related question about the *selective* nature of our causal concept. Peter Unger [1977], for example, argued that the word 'cause' was selective such that background conditions simply did not qualify as 'causes' by definition. Lewis, on the other hand, took his philosophical mission to be one concerned with a *pre-selective* [1973, p.558-559] account of causation—a task aimed at determining the broadest set of causes of an event. For Lewis background conditions are causes, for Unger they are *merely* background conditions. Clearly there are two distinct concepts at play here, one broad and non-discriminatory, the other narrow and discriminatory, and it seems that each different concept maps onto a different treatment of (11). The selective account of 'cause' can treat the oxygen as a background condition, in the right context, whereas the pre-selective account cannot. Accordingly the selective account will accept (11) whilst the pre-selective account will not.

Even if Unger's account turns out to be a good account of the word 'cause' in English, that does not preclude us from questioning what causal concept remains once selection effects have been removed. Since Lewis is explicit that he is considering such a pre-selective notion, he is entitled to reject (11) by fiat, even if not by intuition. I reject

it twice over—once by intuition and once again by fiat. I do not share the intuition that (11) appeals to,[21] but even if I am later shown to be in the linguistic minority, it will not matter to the concept that I am considering, only my entitlement to name that concept 'cause'. Since (11) was all that stood in the way of a pragmatic solution to Selection cases, rejecting it makes the Gricean strategy available once more.

A dialectical aside: Schaffer does not directly consider the **Causal Inquiry** or **Multiple Alternative** cases in his discussion of the prospects of a pragmatic solution. Rather, he focusses on the establishing that no *single* existing pragmatic mechanism could handle all of the cases he presents as data—what may work for **Selection** cases (relevance) will not help with the *Sentential Sensitivities* examples. This is only important if we are committed to a *wholly* pragmatic account of the contextual variation. My CCT view is not. Note, though, the work that Schaffer's false dichotomy is doing here: anyone who does not adopt a wholly pragmatic account is seemingly committed to a semantic treatment in which the contextual variation is traceable to the word 'causes'. Mixed views, or semantic approaches which trace the variation elsewhere are simply overlooked.

So, having denied (11) and explained why some might accept it, I conclude that the CCT view, combined the Gricean maxim of *relation*, can account for all of the data that Schaffer presents in arguing for his preferred contrastivism.

2.5 Contrastivism in General

Thus far I have argued that events have inconstant modalities across contexts and that this insight, coupled with a counterfactual analysis of causation gives intuition-matching results in the relevant test cases of contextually sensitive causal claims. So, the CCT view provides an account of the context variation of causal claims. I have also pointed out that this approach is overlooked in Schaffer's argument in favour of a contrastivist semantics for causal claims. What is not argued here is that the CCT view represents a definite improvement over Schaffer's contrastivism. The treatment of contextual variation cannot be expected to settle that question alone, but here I offer some positive reasons in favour of the CCT view.

First, the central motivation behind the CCT view is the realisation that event expressions invoke different modalities across contexts quite independently of any reference to 'cause'. As such we should expect our event-involving causal claims to exhibit at least some context variation in virtue of the events that constitute them. This provides the CCT view with *independent* motivation, i.e. motivation that is independent from the desire to conform to any particular causal theory.

Second, for Schaffer, all context variation is due to shifts in implicit contrast places—a view which requires a quaternary rendering of the causal relation. It offends parsimony to posit four variables in a theory (c, e, c^* and e^*) where two will do, and according to the CCT view two *will* do, just as common sense would have

[21] As with Schaffer's "more sophisticated speaker" [p.43] I consider it a background condition, but presumably unlike Schaffer's sophisticate I consider background conditions to be causes.

predicted: when we say c caused e, we speak of two things, not four. It seems that CCT may offer a more parsimonious account.

Third, the CCT view reads the counterpart relation from the context just as the contrastivist reads the contrasts from the context. For any given contrast class for c, say c^*, there is a counterpart relation for c such that c^* is just equivalent to $\neg Oc$—one in which the only $\neg Oc$-worlds are those c^*-worlds. If the context justifies a particular contrast class, then we can expect that same contextual information will justify the equivalent counterpart relation. Schaffer does not offer an analysis of causation [2005, p.348] but the contrastive account on offer is supposed to improve on the original Lewis-style accounts despite this precisely because it has the context-sensitive flexibility to consider different contrasts to the c and e events under consideration. If CCT can match this feature, and can do so within a more parsimonious structure, then that is reason to think it might offer a more compelling package overall.

Of course it remains to be seen if such detailed counterpart information can be read-off the context in this way. An account is owed by the defender of CCT.[22] However, I would say that an account is past due from the contrastivist as to how this is to work on their picture, which remains underspecified—Schaffer ends his account in puzzlement having failed to find appropriate link between context and his preferred semantic theory. Even assuming each are able to give a satisfying account of how to read the context, there remains cases where context is almost entirely absent, as in the following adaptation of Hitchcock's example:

13. Susan's stealing the bicycle caused her to be arrested.

In discussing this case earlier, I showed that shifting an emphasis on 'stealing' in (2) to 'bicycle' in (3) signalled a shift in the counterpart relation being invoked. The contrastivist prefers to think of the event as being static throughout, but that a contrast shift has occurred which tracks the shift in emphasis. This way the CCT and contrastive approaches both match intuitive acceptance and rejection of the respective claims.

However in the emphasis-free example given (13) what cause-side contrast is being invoked? Without the emphasis (and absent further context to provide it) there needs to be a very general contrast case which does not privilege any particular feature of the target event. To consider a contrast where Susan does something else with the bike (borrows) would be to act as if 'steals' was emphasised. Or to consider a contrast where Susan steals something else (skis) would be to act as if 'bicycle' was emphasised. It seems that the only justified contrast c^*, absent emphasis or other contextual clues, is one where it is not the case that *Susan steals the bicycle*. Without emphasis or further context c^* simply equates to $\neg Oc$. Once emphasis or other contextual cues are introduced a semantic shift is triggered and both CCT and contrastive accounts can track this shift.

In short, the CCT view is independently motivated, parsimonious and handles the standard test cases at least as well as the contrastivist alternative from Schaffer. That

[22] I offer the beginnings of such an account in Chapters 3 and 4.

it can also be expected to match the contrastivist treatment in general suggests that it should at least be considered a viable alternative.

2.6 Conclusion

In this chapter I have argued for a counterpart-theoretic treatment of events. Whilst the actual-world referent of an event expression can remain fixed across contexts and re-descriptions, the counterpart relations that the expression invokes may vary. I argued that this counterpart-theoretic view bridges the gap between fine-grained and coarse-grained accounts of event ontology—it has the precision of the first with the parsimonious ontology of the second. I have demonstrated that context variation is evident in our inconstant *de re* modal attributions concerning events and I have shown that this in turn impacts on certain counterfactual conditionals involving those events (§2.1).

On the assumption that the truth of such counterfactual conditionals can be indicative of causal connection, then tracking shifts in the counterpart relations that events fall under can explain why the acceptability of certain causal claims varies with context (§2.3, §2.4): the acceptability of the claim varies because its truth value varies. Contra Schaffer, I propose that the context sensitivity in causal claims is traceable to the counterpart variation of events across contexts, not to the presence of the word 'causes' (§2.5).

If this is correct, then the CCT view that I have proposed—coupling counterpart theory and a counterfactual test for causation—retains a parsimonious binary model of causation and accounts for the variation of causal claims across contexts. It should therefore be considered a viable, even attractive, alternative to the contrastivist contextualism offered by Schaffer.

3

Dare to Be a Doctor

Counterfactual theories of causation are beset by so-called pre-emption counterexamples. In such examples there are two candidate causes which suffice for an effect and so the effect *depends* on neither, and yet intuition is clear that one of these candidates is the cause and the other is not. Standard counterfactual analyses, and contrastive theories too, seem to give the wrong result in a particular species of pre-emption case known as late pre-emption. In this chapter I aim to show that the CCT view that I introduced in Chapter 2, coupled with a plausible pragmatics of our causal talk, can give intuition-matching results in cases of late pre-emption.

I have argued that by adopting a counterpart theoretic view of events we could account for certain instances of context-variation in our causal attributions. This was in the service of my first aim in this thesis: (I) to give an account of our everyday causal talk. Building on this view, I will give an argument in support of a *fragile* view of events—a view by which even minute alterations in the timing or manner of an event bring about a new event. At one point Lewis complained that the project of mapping out a defensible account of event fragility was 'not so much unfinished as unbegun' [1986b, p199]. I take this chapter, and the thesis on the whole, to be the beginnings of such a project. I will argue that the much-repeated 'who would dare be a doctor?' riposte from Lewis fails as an argument against the fragilicist.

My treatment will prompt a worry about the apparently fluctuating standard of fragility to be applied. I will identify a key asymmetry in our attribution of event fragility and use this asymmetry to show that a principled, and well-precedented, set of pragmatic maxims can be deployed to achieve intuition-matching results, even in cases of late pre-emption. The key is accepting that events *can* be fragile, not that they always are.

3.1 Counterfactual Dependence and Pre-emption

Hume argued that we never directly perceive the causal connection between occurrences. When c causes e, it is c and e that we are acquainted with, not the 'causing' that apparently links them.[1] According to Hume the best we could claim was that one thing causes another if and only if those sort of things, across a multitude of cases, occur with just the right sort of regularity. According to Lewis [1973], the question of what the right sort of regularity was had dominated the philosophy of causation for over 200 years without resulting in a satisfying consensus.

A recalcitrant problem for such a view concerns cases of epiphenomena. If one cause gives rise to two effects then the effects will stand in just the same regular relation with one another as they do to the putative cause. So, low air pressure causes low barometer readings and storms but, even if they were to occur with perfect regularity, low barometer readings do not cause storms. Hume's analysis is obliged to say that they do and that is a failure of the analysis.

However, Lewis spotted greater potential in a rather different offering from Hume: 'if the first object had not been, the second never had existed' [1975, p.76]. In other words, without the cause the effect would not have taken place—a strikingly different proposal to the one it is offered in support of. This dependence of the effect upon the cause goes beyond a simple actual-world regularity and speaks instead of other-worldly contingency. This counter-to-fact reasoning about what would have happened in the absence of the cause seems to be just what is missing in the air pressure problem case—absent the low air pressure, the storm would not take place and so the low air pressure causes the storm, but absent the barometer reading (and ceterus paribus) the storm would still occur and so the barometer reading does not cause the storm. This looks like a brighter prospect in terms of tracking our causal ascriptions.

Thus, Lewis offered us the following analysis of causation: c is a cause of e if and only if c and e are linked by a chain of causal dependence, where causal dependence is analysed as counterfactual dependence between distinct events. Event c counterfactually depends on event e iff were it not the case that c occurred, it would not have been the case that e occurred. More formally:

$$\neg Oc \mathrel{\Box\!\!\rightarrow} \neg Oe$$

The elegance of the analysis belies a complicated issue—how do we tell which would-be claims are true? In other words, how do we establish the truth conditions of the counterfactual conditional? As of 1973, Lewis [2001] and Stalnaker [1968] had each offered a broadly similar semantics of counterfactual claims which, they claimed, established objective truth conditions for the $\Box\!\!\rightarrow$ operator. Roughly speaking, $\neg Oc \mathrel{\Box\!\!\rightarrow} \neg Oe$ is true if and only if, in all of the closest worlds where c does not occur, e does not occur.[2] Here closeness is to be understood in terms of overall similarity to the actual world such that closer worlds are more similar to the actual world than more distant

[1] For a critical discussion of this typical reading of Hume and its implications, see Beebee [2009].

[2] As previously mentioned, I am using a simplification which implies the Limit Assumption. I believe this is harmless. See Lewis [1973, p.561].

worlds. Needless to say, this similarity ranking of worlds is controversial and Lewis left it intentionally under-specified.

However, the benefits to be accrued from thinking in counterfactual terms, especially in causation, justify the effort that is required to understand the similarity criterion. As interesting as it is I won't discuss that issue in this thesis other than to accept that the counterfactualist owes a fuller account of the specifics here.[3] Here I only wish to point to three key features of Lewis's analysis: First the dependence (or chains thereof) that holds between the effect and the cause is central to the analysis—if this dependence fails to hold in a case of causation then the analysis fails; Second, the relata are events. That is, the c and e in the causal expression 'c caused e' are to be understood as events. Indeed all cases of genuine causation should be understood as cases of relations between events; Third, a given event is taken to occur in more than one world.

This third feature of Lewis's analysis was later [1986b] clarified by Lewis as meaning that events were genuinely transworld entities. In the preceding chapter I argued that there was a benefit to understanding the context-sensitive nature of our causal ascriptions if we adopt a counterpart theoretic view of events instead. Strictly speaking, both theories of event modality are compatible with Lewis's original counterfactual analysis. In this chapter, however, I will talk within the counterpart-theoretic framework as I believe it allows us to clearly express, and therefore disentangle, important ambiguities in our causal expressions.

On a counterpart-theoretic view, events do not literally occur in many worlds (they are concrete individuals which only exist in a given world) but we can nevertheless think of events in other worlds as being counterparts of events in our world. So when we say that the storm occurs in many worlds, we mean that there are many worlds in which there is a counterpart of the storm. When we talk of worlds where the storm does not occur, we mean that those worlds contain no counterpart of the actual world storm. The counterpart relation that applies in the context of a given causal claim—i.e. the relation that fixes what counts as a counterpart for each of the individuals being referred to—is partly determined by the context and, when the individuals are represented in a sentence, the mode of presentation of that causal claim.

With this in mind, I turn to the problem of pre-emption.

3.1.1 Dependence and Guarantee

Suppose that you depend upon your wages to pay your rent. If you later come to be supported by a wealthy benefactor who offers to guarantee your rent in the event that your wages fail to arrive, then you no longer depend upon your wages to pay your rent. You still use the money that arrives from your wages to pay your rent so the internal mechanics of the transaction are the same, and you may never call on the guarantee so that it remains unused, but nonetheless the dependence relation is undermined by the presence of your guarantor. The lesson seems to generalise—dependence relations can be undermined by the presence of a guarantor or back-up.

[3] For an interesting short discussion, see Schaffer [2004b].

Applying this lesson to Lewis's analysis gives rise to the problem of pre-emption. The pre-emption class of objections is built upon the guarantor structure so that an effect is guaranteed in such a way that it no longer depends upon the putative cause. Such examples appear to refute Lewis as they are cases where we have causation but no dependence, a result that his analysis rules out. So much the worse for that analysis, many have thought.

Pre-emption cases come in a variety of forms: early pre-emption, late pre-emption, trumping pre-emption, pre-emptive prevention and super pre-emption. The guarantor structure is common to each, but the cases vary in the details. In this chapter I focus on the cases Lewis seemed to find most troubling—late pre-emption. I will return to each of the others in subsequent chapters.

> **LP:** Billy and Suzy are out to vandalise. Each throws their own rock accurately at a window but Suzy throws faster and her rock reaches the window first. The window breaks and Billy's rock sails through the void.

Whilst Billy's rock does not connect with the window, it nonetheless guarantees that the breaking of the window will occur. As such, there is no counterfactual dependence of the window break upon Suzy: absent Suzy's throw, the window still breaks. So, whilst it is plain that Suzy caused the window to break, the counterfactual analysis says that it does not. For Lewis, this was a terminal failure of his 1973 analysis.[4]

Yet it seems obvious that the window breaking that would occur absent Suzy is a different breaking—it would happen later, and presumably in a different way. That suggests that there are two window breakings under consideration: the one that comes to pass when Suzy throws and the one where she doesn't and Billy's rock strikes instead (call them s and b for short). If s and b are different, then there is no one event which is caused by Suzy and backed-up by Billy. Rather there are two events, s which counterfactually depends on Suzy and b which does not. If this is the case then s is the actual world event we wanted to know about, and s depends on Suzy. As such, the counterfactual analysis gets the case right—Suzy causes s. Lewis characterises the solution thus:

> There is an obvious solution to cases of late pre-emption. Doubtless you have been waiting impatiently for it. Without Suzy's pre-empting rock, the [window] would still have shattered, thanks to Billy's pre-empted rock. But this would have been a different shattering. It would, for instance, have happened a little later. The effect that actually occurred did depend on Suzy's throw. It did not likewise depend on Billy's. Sometimes this solution is just right and nothing more need be said. [Lewis, 2004a, p.85]

However, Lewis rejected this as a general response to late pre-emption cases. In the next section I will expand on Lewis's reasoning and argue that we should not be convinced by a well known argument he gave against fragility, but that we should respond to a related problem.

[4] In his [2004a] Lewis offers a revised analysis on the basis that his earlier analyses failed to adequately handle pre-emption.

3.1.2 Who Would Dare be a Doctor?

An event is fragile if and to the extent that alterations in timing or manner make it a new event. Event s is fragile in this sense, since if the window had broken a little later it would have been a different event, b, instead. The response which distinguishes s and b, and thereby rescues the counterfactual analysis in the LP case above, requires that we adopt a fragile view of events.

Lewis thought that fragility was deeply problematic and he offered the following argument: if every alteration in the timing or manner of an event rendered it a different event then every alteration of the past that in any way changes that event, will have caused it. Surely not! If that were the case then every intervention on a patient by a doctor that even slightly altered the timing and manner of that patients death would have caused the death. On this account every doctor kills every patient. On those terms, 'who would dare be a doctor?' [Lewis, 1986b, p.250].

In other words, too many things that we do not ordinarily deem to be causes, are causes under the fragile view. The effect event is too counterfactually sensitive to alterations to track our intuitive causal ascriptions. By the time we distinguish s from b we have set a precedent whereby all manner of paradigmatic non-causes must be considered causes under the counterfactual analysis. A dog barking several streets away creates minor vibrations in the window just at the point of its breaking. On a fragile view, the dog barking caused the window to break. Lewis took this to complete a *reductio* of the fragile position [1986d, p.198].

Before moving to respond, it is important to make clear what is being disagreed about here. On the one hand we have the view that the window breaking event could have happened later, or differently. On the other hand we have the view that *that* window breaking event could not have happened later or differently. The views are differing on what features of the event are essential, what range of counterparts the event is taken to have.

One sort of *robust view* of the event takes it to be essentially a window breaking, and only accidentally at that time and in that way. As such all manner of other-world window breakings would have the same essential features and therefore qualify as counterparts of the window breaking event in our world even where the accidental features vary. By contrast, a *fragile view* of the event takes it to have many more essential features and far fewer accidental features than the robust view does. Only a restricted set of other worlds contain events which have all of those essential features and so the event has relatively few counterparts. It is often useful to refer to the event itself as robust or fragile when it is taken to have a relatively large or small number of counterparts respectively. However, it should be made clear that it is not the event itself that is robust or fragile. The event is simply a region of a world and the robustness or fragility of that event is a comparative measure of the number of counterparts the event is taken to have in a given context. This idea is introduced and defended at length in Chapter 2.

So, it is the counterpart relation that is at issue when considering the 'two events' response to late pre-emption problems. Throughout his career Lewis changed his view on which features of a thing were essential and which were accidental. In 1968 he

argued that (most) objects had fixed essences and therefore had a determinable set of counterparts. By the time he wrote Things Qua Truthmakers [2003] Lewis had conceded that the counterpart relation was probably not fixed and it was therefore a vague and context dependent matter which other-worldly objects were counterparts.[5] This later view supports the central claim of my thesis regarding causal *talk*: there is no mind-independent fact of the matter about which counterpart relation applies to events when we assess the content of an ordinary causal claim.

So, there is no independent fact of the matter about how fragile or robust the window breaking event is when we say 'Suzy caused the window to break'. Both the robust and fragile views remain available and so the talk of the window breaking event remains ambiguous across these two readings. In fact, it remains ambiguous across an entire spectrum of readings.[6] Where an event sits on this spectrum dictates what we might refer to as the *scope* of the event—the range of counterparts that the event has in a given context. Interestingly, data about what events are considered outside the range of counterparts is evidence of what position on the spectrum is being considered since this partially delimits the scope of the event.

This delimiting feature is exploited by causal contrastivism in an illuminating way. Causal contrastivists come in a variety of forms (Maslen [2004a], Schaffer [2005, 2012a], Northcott [2007] and List & Menzies [2010]) but one common feature is the following insight—we can disambiguate a causal claim by rendering it in a contrastive locution. In Chapter 2 I introduced Schaffer's version of this theory which takes both the cause side and the effect side to be contrastive, though often implicitly so. Making that contrast explicit involves stating clearly what alternative is being considered in the form: c *rather than* c^*. Take the following causal claim as an example: 'moderate smoking causes cancer'. This claim is ambiguous across two readings: (i) moderate smoking *rather than* heavy smoking causes cancer, (ii) moderate smoking *rather than* non-smoking causes cancer. The first claim seems false, and the second true, since moderate smoking is likely to reduce your chances of cancer relative to heavy smoking, but increase your chances relative to non-smoking. The contrastive 'rather than' clause disambiguates the claim but it is important to note that this contrastive locution simply gives us a single reference point for what is *not* to be considered a case of c occurring. This makes it a useful short cut in ordinary language for specifying a limit on what counts as c occurring (by giving an example of what does not count). So, whilst I maintain that a counterpart theoretic understanding of the relata is preferable to a contrastivist one, I will use the contrastive locution as a short cut instead of fully specifying what qualifies as a counterpart of c.

Applied in the case of late pre-emption we can see the difference between the robust and fragile readings of the window-break event more clearly when they are rendered in contrastive language.

[5]For a detailed discussion of Lewis's trajectory, which I substantially simplify here, see Beebee & MacBride [2014].

[6]The robust and fragile views do not even represent the extremes of the spectrum—at one end the event has every other region of every other world as a counterpart, at the other it has no counterpart other than itself.

Robust1 Suzy's throwing the rock rather than dropping it, caused the window to break rather than not break.

Fragile1 Suzy's throwing the rock rather than dropping it, caused the window to break the way it did rather than break a little later.

With Billy acting as guarantor, Robust1 comes out false—even if Suzy had dropped the rock, the window would still have broken. However, Fragile1 comes out true since if Suzy had dropped the rock, the window would have broken a little later. Suzy's throw does not make a difference between the window breaking and not breaking, but rather it makes a difference between the window breaking the way it did and the window breaking a little later.

When the contrastive interpretation is applied to Lewis's dog bark case, we can see the equivalent two readings:

Robust2 Dog's barking rather than not barking, caused the window to break rather than not break.

Fragile2 Dog's barking rather than not barking, caused the window to break the way it did rather than break very slightly differently.

Again, Robust2 is false because Billy and Suzy guarantee that the window will break—Dog's bark does not make the difference between breaking and not breaking. Fragile2 is true, however, since *ex hypothesi* Dog's barking does alter the window break very slightly—Dog's bark does make the difference between the window breaking the way it did versus it breaking very slightly differently.

Finally, returning to the Doctor case, the contrastive interpretation yields the following:

Robust 3 Doctor's intervening rather than not intervening, caused the patient to die rather than not die.

Fragile 3 Doctor's intervening rather than not intervening, caused the patient to die one way rather than die another.

No factor can ever be the difference between someone dying and not dying at all, since it is (nomologically) impossible for a human to not die eventually. So, the doctor never had any chance of influencing that. All we can ever do is alter the how and the when. Fragile3 is the only sensible reading of such a claim—*ex hypothesi* the doctor's intervention made some difference to the timing or manner of death, therefore the doctor's intervention was the difference between the patient dying the way that they eventually do and them dying some other way.

Lewis's retort of 'who would dare be a doctor?' draws on the intuitive falsity of the idea that every doctor kills their patient. But it could well be the case that every doctor alters the timing and manner of their patient's death without that warranting the claim that this meant they had 'killed' the patient in question. Lewis seems sensitive to this point elsewhere when he argues that your birth is in fact a cause of your death, despite

the intuitive oddity of the claim [2004a, p.101]. The point is that 'killing' requires more than simply having some impact on the timing and manner of the death, it requires some particular difference is made by the act in question. That requirement is not satisfied simply by establishing that a doctor alters (perhaps for the better) the timing and manner of the death in question. On the assumption that any difference maker deserves to be considered among the many genuine 'causes' of an effect, as Lewis's 'broad and non-discriminatory' [1973, p556] analysis suggests, then the doctor must be allowed to be a cause. On the further assumption that the doctor does not kill every patient we should not conclude that the causal analysis is wrong but rather that being one among the many causes of a death does not automatically render you a killer.

The claim that every doctor kills every patient is absurd, just as Lewis says it is, but there is no reading of events which is committed to this claim. The Robust3 reading of the death event would make it the case that if the doctor caused the death then the doctor would have killed the patient. However the Robust3 reading also denies that that every doctor counts as a cause of the death merely by altering the timing and the manner of that death. The Fragile3 reading of the death event does consider every doctor to have causally contributed to the death in virtue of (even slightly) altering the timing and manner of the death and thereby every doctor counts as a cause of every one of their patients' deaths. However the fragilicist would not be committed to the idea that mere causal contribution is enough to make you a killer since the causal contribution in question need not have made the difference between life and death, but could merely have altered the how and the when. So, either every doctor is a cause of every death, but being a cause is not enough to make you a killer (Fragile3), or being a cause *is* enough to make you a killer, but not every doctor is a cause of every death (Robust3). Lewis is conflating the counting of causes as endorsed by a fragile reading with the import of being a cause on a robust reading. Lewis's argument rests on an equivocation.

However, there is a more sophisticated argument lurking just under the surface. I said that the fragilicist was not committed to the absurd result in Lewis's doctor argument, but that is not to say that a uniformly strict standard of fragility of events is an attractive view of events. Such a view commits us to accepting that the gravitational pull of Jupiter was a cause of my drinking the coffee before me and that Billy was a cause of the window breaking since his gravitational influence made some difference to the effect event. Such claims are counter-intuitive in light of the context we began with: the context of kids throwing rocks at a window. However, to conclude that fragility is the problem is premature. Looking back at the window break case we can see that the window would likely have broken eventually at some point in the future, by other vandals, by demolition or by some apocalypse. We were never worried about those scenarios because there was already some implicit limit on the extent of the window breaking event, even on the more robust renderings being considered. In other words all the readings are fragile to some extent so it cannot be fragility itself which is the problem. We can accept that there will be some contexts in which such minute contributions as Jupiter and Billy's gravity have to their respective effects may be salient, but not in any normal context where some vandalism is being investigated. So there is a contextual element here that the constant fragilicist ignores to their

detriment. What is needed is an inconstant or, better, *flexible* standard of fragility which remains sensitive to context.

The more sophisticated worry we might have about such a flexible standard of fragility is that whilst events surely *can* be fragile, what is to say that the context of the pre-emption cases is such as to justify considering them to be fragile in those circumstances. What is it about pre-emption cases that justifies treating the event as fragile to just the *right* degree so as to get the right result (that Suzy is a cause of the window breaking), but not the wrong results (that the dog's bark and Billy were too). In the next section I will defend a view of the pragmatics of causal discourse which meets the challenge. However, before moving on I should make clear the role that contrastivism played here.

By plugging in the 'rather than' clause, the contrastive locution makes explicit what, in a given context, is *not* to be considered a counterpart to the c or e events. c^* and e^* do not pick out every variety of $\neg Oc$ and $\neg Oe$ but some specific case of $\neg Oc$ or $\neg Oe$ (*drop the rock* and *break a moment later* respectively). For it to make sense to contrast c and c^* requires that c^* falls outwith the scope of Oc in that context. Of course, finding what falls outwith some scope does not serve to completely define that scope, but it does at least partially delimit it. For example 'the window breaking when it did rather than a minute later' rules out counterparts of 'window breaking' that occur a minute later. It likely also rules out window breakings two minutes later, ten minutes later and an hour later. The contrast event c^* implies a limit on the counterpart relation of the target event c (and therefore a limit on the scope of Oc-worlds).

Note that the benefit of restating the claims in a contrastive form could have been attained by explicitly delimiting the counterpart relation of the target event from the outset. The contrast event signals the limits on the counterpart relation of the target event, but if the target event and its counterpart relation had been thoroughly specified to begin with, then that signal would be redundant. It is nevertheless a very useful short cut to specifying the relevant portion of the counterpart relation in ordinary language.

I make this point here to clarify that a contrastive semantics, and the baggage it carries, is not strictly necessary to make my point about Lewis's argument.[7] I think I have good reasons to avoid commitment to the contrastivist program, and I will come to discuss them later in this thesis, but I still think that the contrastive form is an excellent tool for disambiguating our event reference in causal claims—the window breaking example being a case in point.

3.2 Pragmatic Maxims

In Chapter 2 I argued that there is an important variable in our causal claims, one that can alter the truth or falsity of the claim: the modality of the events in question (captured by a counterpart relation). In the first section of this chapter I argued that this variable is at the heart of the debate about so-called fragile events. In this section I

[7]Recall from Chapter 2 that the contrastivist needs to posit four causal relata rather than two. I consider that ontological baggage.

will propose certain rules or maxims of accommodation for our causal discourse which, when combined with the observations of the previous section, offer a principled reading of late pre-emption cases such that the counterfactual analysis gets the cases right. I will start with quite general and well-precedented pragmatic maxims, then show that the pragmatic variables in causal claims have an important asymmetrical feature. This will allow me to derive a more specific set of maxims for causal claims and show how they apply to pre-emption examples.

3.2.1 Accommodation and Asymmetry

In ordinary discourse it is standard to interpret utterances that are not entirely explicit, or carry some unstated implication, in line with certain rules or maxims. Famously, Grice proposed a governing principle of well-conducted discourse—the Cooperation Principle:

> Make your conversational contribution such as is required, at the stage at which it occurs, by the accepted purpose or direction of the talk exchange in which you are engaged. [Grice, 1968, p.307]

Grice proposes that we interpret what people say as if they are acting upon this general principle. He went on to specify more detailed sub-principles and I will invoke some of these below. Lewis [1983e] argued in a similar vein that we should aim to be accommodating of what people say so as to most charitably interpret what they said. In particular Lewis offered this rough *rule of accommodation*:

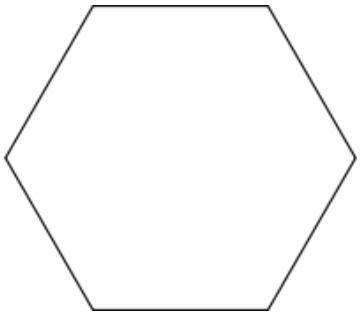
> If at time t something is said which requires presupposition P to be acceptable, and if P is not presupposed just before t then—*ceteris paribus* and within certain limits—presupposition P comes into existence at t. [Lewis, 1983e, p234]

By accommodating fellow speakers in this way we are licensed to shift the implicit presuppositions or implicature charitably. Before I discuss the application of this rule to causal discourse, let me first introduce a useful example of context-relative vagueness from Austin:

1. France is hexagonal.

There will be contexts in which 1 is acceptable, for example when teaching a child how to remember which country on a map is France, and there will be contexts where it will not, such as providing a diagram of a hexagon for geometry class. Given such contextual variation, it is not possible to give a once-and-for-all standard of tolerance for the concept 'hexagonal'.

However there remains an important asymmetry of precision, someone who accepts that France is hexagonal had better also accept:

2. This shape is hexagonal. 

And yet it remains open to someone else to insist that the slightly jagged lines disqualify it from being a hexagon. Such a person is adopting a very strict standard for hexagonality and so should also reject 1 if they are to be consistent. I take this asymmetry to be general regarding precise and imprecise claims.

Adopting some more accommodating interpretation of what is said is advocated by both Grice and Lewis but in cases of asymmetrical entailment, such as the cases of precision above, a further restriction is needed. Suppose that there are three standards of precision concerning what it is to be a hexagon: the strictest standard where only an absolutely perfect hexagon will do; a less strict standard where hand-drawn shapes and imperfect prints will suffice; and a much weaker standard where anything even broadly of the right shape (e.g. France) can be considered hexagonal. To make virtually every claim of hexagonality true, we could simply take the speaker to presuppose the weakest standard every time. To make every denial of hexagonality true we could simply take the speaker to be presupposing the strictest standard on every occasion. However, the person who claims that the hand-drawn shape is hexagonal need not be committed to the idea that France is hexagonal, so if we take the speaker to presuppose the weakest standard of hexagonality we commit them to a position that they may not endorse. Similarly with the person who wishes to deny that France is hexagonal: they need not be committed to denying that any hand-drawn shapes are hexagonal. The important point here is that so far Lewis's rule of accommodation only tells us to impute *some* presupposition that would make the claim true. Yet there is some further principle required in order that the presupposition adopted does not overcommit the speaker on related claims. This thought is related to Grice's more refined maxims of Quantity [1968, p.308]:

3. Make your contribution as informative as is required (for the current purposes of the exchange).

4. Do not make your contribution more informative than is required.

These maxims concern the level of information that a speaker ought to produce, but crucially they say that the speaker should go far enough, but no further than is required. Applying this more general lesson to the attribution of presuppositions to a speaker, we should accommodate as strong a presupposition as is required, but no stronger than is required, to make the claim true. I use 'as strong' advisedly here in place of 'as many' since I want to highlight the fact that it is not just the number of presuppositions that we attribute that should be subject to the rule but also, as

in the case of precision, the *strength* of the presupposition. When we attribute a presupposition to the speaker who denies that France is hexagonal, we should suppose some stricter standard for 'hexagonal' but no more strict a standard than is required. Only certain presuppositions will have this asymmetrical scale-like feature where there is some dimension, and direction, of extension which needs to be considered.

In the next section I will show that the context variable component of causal claims—the counterpart relations of their constitutive events—demonstrate this feature of extent. I will then go on to derive some more specific maxims for causal discourse and apply these to the problem cases of late pre-emption.

3.2.2 Tend to Fragile, Tend to Truth

In this section I will argue that a causal claim which is false when the putative effect is taken to be robust, can still be true when the putative effect is taken to be fragile. This will justify tending to a fragile reading of the effect under consideration in ordinary causal claims. I will first argue that all true causal claims with a robust effect are true when the effect is *strictly* more fragile.

Here the terms 'fragile' and 'robust' should be clarified. An event is fragile if, and to the extent that, varying the timing or manner of the event renders it a different event. An event is robust if, and to the extent that, varying the timing and manner of the event do not render it a new event. In counterpart-theoretic terms, a fragile event is an event taken to have *relatively* few counterparts, and a robust event is an event taken to have *relatively* many counterparts. Of course the event itself is not robust or fragile on my view but describing an event as robust or fragile indicates that the event is taken to have relatively more or relatively less counterparts respectively in each case.

To indicate that an event c or e falls under a counterpart relation x, I will represent c in the context which invokes that counterpart relation by writing c_x or e_x. I now introduce the further notion of an event being *strictly* more fragile under one particular counterpart relation ($< n$) than another ($n >$) (i) when every counterpart of the event under $< n$ is a counterpart of the event under $n >$; and (ii) when the reverse is not the case. In other words, every essential feature of the event when it is taken to be robust is an essential feature of the event when it is taken to be fragile but not vice versa. For the purposes of notation I will express the relative fragility of the event by signifying its counterpart relation relative to n: $e_{<n}$ indicates that e is to be taken to be strictly more fragile than when under counterpart relation n (and I will use $<< n$ and $<<< n$... for the progression of counterpart relations that render e progressively more strictly fragile than $e_{<n}$); whereas $e_{n>}$ indicates that e is to be taken to be strictly more robust than when under counterpart relation n (and I will use $n >>$ and $n >>>$... for the progression of counterpart relations that render e progressively more strictly robust than $e_{n>}$).

We can now make the following claims:

5. Any world in which there is a counterpart of e under relation $< n$ will also have a counterpart of e under relation n (by (ii) above).

And, since every $e_{<n}$-world is an e_n-world:

6. Any world in which there is no counterpart of e under relation n can have no counterpart of e under relation $< n$, i.e. $\neg Oe_n$ entails $\neg Oe_{<n}$.

Applying this to a counterfactual account of causation yields an interesting result. Once we fix which worlds are $\neg Oc$-worlds (i.e. once we fix the reference of c and its counterpart relation n) then if all of those worlds are $\neg Oe_n$-worlds then, by 6 above, all of those worlds are also $\neg Oe_{<n}$-worlds.[8] More formally:

7. $\neg Oc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_n$ entails $\neg Oc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_{<n}$.

This means that any true causal claim of the form $\neg Oc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_n$ entails the truth of the claim $\neg Oc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_{<n}$. In other words, if a causal claim is true when e has a given counterpart relation then that claim is true for any strictly more fragile counterpart relation for e. An example may help bring this out: on the assumption that the window broke into 357 pieces, if it true to say that if the rock throw had not occurred ($\neg Oc_n$) the window would not have broken ($\neg Oe_n$), it must also be true to say that if the rock throw had not occurred ($\neg Oc_n$), the window would not have broken into 357 pieces ($\neg Oe_{<n}$).

The reverse is not true. $\neg Oe_{<n}$ does not entail $\neg Oe_n$. For example, for the window not to have broken in 357 pieces does not entail that the window didn't break some other way. So if a causal claim is true at some more fragile standard for e, that does not guarantee that it will be true at any more robust standard for e. This means that it is possible for a causal claim concerning an effect to be false under some robust reading but for some claim, concerning the same events, to be true under some more fragile reading of the effect. McEnroe's serve has all of the same counterparts as McEnroe's awkward serve but it has many more besides, for example, the graceful ones. This makes the awkward serve strictly more fragile than the serve. Even though it is false to say that McEnroe's tension caused him to serve, it is still true to say that his tension caused him to serve awkwardly.

I have talked here of the features of the event but not its timing, and yet in standard cases where fragility is discussed it is the *timing* of the event (rather than the manner) that is considered. Is it the same death if it happens a moment later? A day later? A week? The period in which the event could have been taken to occur in part dictates which counterparts the event is taken to have in that context, and so is one measure of the fragility of that event. If A knows that an event had to have occurred within a minute then A will attribute a counterpart relation for that event that includes counterparts within the minute span but none outwith it. If B only knows that the event

[8] This notation is helpful but potentially misleading. I take it that a given counterpart relation relates c to its counterparts and e to its counterparts. That counterpart relation is a function of all manner of contextual parameters but the resultant relation is one that relates all of the individuals (c, d, e, $f\ldots$) in the situation to all of their counterparts. However I am trying to discuss the impact of shifting the counterpart relation in respect of only one individual in the environment: e. To signify that I am considering relations that assign the same counterparts for c, d, $f\ldots$ but different counterparts for e, I am only altering the subscript for e. That does not imply that c and e are falling under *different* counterpart relations but rather that c has the same counterparts under two candidate relations and e does not.

occurred within a given hour then they will consider the event to have counterparts within that range as well as counterparts within the minute that A knows the event occurred in. As long as A and B agree about the essential features of the event besides its timing, then A's counterpart relation for the event will be strictly more fragile that B's.[9] Time is just another dimension of fragility.

So these two observations tell us that a true causal claim is never made false by treating the effect event as more strictly fragile (as in the window-break example), but a false causal claim can be made true by doing so (as in the McEnroe example). This means that in cases where there is ambiguity or doubt regarding the counterpart relation being invoked in a given context, if we tend to interpreting the effect as being more fragile, we will tend to make the causal claim in question true.

3.2.3 Maxims of Causal Discourse

So far I have shown that some very general and well-precedented pragmatic accounts of our general discourse advocate a charitable reading of what is said. I have argued that there is a requirement for a limiting maxim on those cases where the extent of a given presupposition is open to interpretation. I have also argued that the extent of the counterpart relation attributed to events in causal claims is important to the truth of those claims and, along with the example of precision discussed earlier, interpretations of the counterpart relations at play in a conversation have an important asymmetry. Here I aim to bring these elements together in establishing certain minimal maxims for interpreting causal discourse.

Lewis's revised version of his rule of accommodation will be useful here:

> If at time t something is said that requires component s_n of conversational score to have a value in the range r if what is said is to be true or otherwise acceptable; and if s_n does not have value in the range r just before t; and if such-and-such other conditions hold; then at t the score-component s_n takes some value in the range r. [Lewis, 1983e, p.240]

This refinement covers not just a static set of presuppositions, but rather of a dynamic and evolving 'conversational score' where the presuppositions shift throughout the language game. Within the conversation are elements (events) which take on a value (counterpart relation) within a range (more/less fragile). This rule of accommodation urges us to take the value to be such that it makes the claim true.

Bearing these notions in mind, consider again McEnroe's serve. Suppose that two attentive observers watch the serve, share a glance and a grimace and then one says: "It was his tension that caused that". The second 'that' here is ambiguous. At the very least it could mean either that the tension caused the serve or that the tension

[9] An interesting complication: there may be some events that occur at intervals—like the arrival of a train every quarter hour—so that there is not a smooth range of times that it could have occurred but rather a set of discrete times. At which point we would better express the occurrences as occurring at one of a disjunction of times rather than within a period of time. Such cases will be rare but they do not impede the point made: one event is strictly more fragile than another if all of the times that the first could have occurred are times when the second could have.

caused the *awkward* serve. Assuming that the person spoke truly, and taking the counterpart parameter of the conversational score to be such that truth would require, we should take it that the counterpart relation is equivalent to that invoked by talk of the 'awkward serve'.

We got to this conclusion by operating on what could be called a 'super-maxim': interpret the utterance as required to make it true. However, by drawing from the observation in the previous section, perhaps we can formulate a more specific maxim for the interpretation of vagueness in causal claims:

Maxim 1 *Take the effect event to be as fragile as is required to render the causal claim true.*

In the McEnroe case, though, I only considered two options for the interpretation of *that*. In fact a spectrum of interpretations were available from the very broad 'made a movement' or 'served' to the highly precise 'served awkwardly, facing east, wearing white shorts and with a heart rate of 165 bpm'. The broad interpretations imply broad counterpart relations and generate false causal claims when a counterfactual analysis is applied—his tension did not cause him to make a movement or serve. These interpretations are therefore ruled out if we are aiming to interpret the claim such that it is true. The highly specific interpretation, whilst providing a true counterfactual claim (and therefore being a value in range r), appears over-specified by normal standards. This is where we should impose a limit on the extent of our supposition, as inspired by Grice's maxims of Quantity. In terms of our counterpart relations, this means that we should not interpret the effect as being so highly specified (fragile) if it is not required. Thus we add another Maxim as follows:

Maxim 2 *Take the effect event to be no more fragile than is required to render the causal claim true.*

However there was another problem with the highly precise interpretation invoked: it wasn't just too precise but it was *implausibly* precise given what we can expect the speaker to know. Taking inspiration from Grice once again, especially his second maxim of Quality, we should not suppose that the speaker was making a claim for which they lacked adequate evidence [1968, p.308]. Presuming that there was no public reading of McEnroe's heart rate at the time of his serve, there is no reason to think that the speaker would know that it was at 165bpm. This bars us from interpreting the speaker's comment in a way that specifies the heart rate. The more general point here is that the epistemic status of the speaker should limit the extent to which we can shift the interpretation of what they say when we are being accommodating or cooperative. In causal terms this means that the extent to which we interpret the effect as being more fragile, is limited to within an epistemically accessible range.[10]

Maxim 3 *Take the effect event to be no more fragile than the speaker could discern or infer at the time.*

[10]Since Maxim 3 limits the range of interpretation, Maxim 3 trumps Maxim 1.

These three maxims are specific to causal claims but are derived from much more general, and well-established, pragmatic maxims. Armed with these maxims, I will shortly turn to the case of late pre-emption and show how these maxims of causal talk can help resolve such problems for counterfactual analyses of causation.

A point of methodology deserves mention before I proceed, however. I have moved smoothly from the maxims which speak of interpreting contributions as though they are true, to interpreting those contributions as though they were true *on a counterfactual analysis*. One might reasonably worry if this step isn't problematically circular, given that a counterfactual account is what is being defended overall. I think it would be an unfair accusation, here, however. I am trying to show that some counterfactual analysis (tentatively the CCT version I introduced in Chapter 2) coupled with some general pragmatic maxims can give a coherent account of our causal discourse. In order to show that it can, it is reasonable to hypothesise that such an analysis tracks truth in order to show that it *can* fit within a workable theory.

3.3 Late Pre-emption

Recall the problematic case of late pre-emption (LP):

> **LP:** Billy and Suzy are out to vandalise. Each throws their own rock accurately at a window but Suzy throws faster and her rock reaches the window first. The window breaks and Billy's rock sails through the void.

Such cases represent a problem for a counterfactual analysis of causation because there appears to be none of the required counterfactual dependence between the effect and, the obvious cause, Suzy's throw. There is counterfactual dependence of the effect upon Suzy's throw if the effect is taken to be suitably fragile but if the effect is super-fragile then it also depends upon a dog barking several streets away. If all events are super-fragile, then every pre-empted thrower, and even every innocent bystander, is a cause of the window's breaking. A workable theory of counterfactual causation needs to be able to give the right answer as to what causes the window to break without giving the wrong answer in a host of new cases.

3.3.1 Fragility as Counterpart Variation

Both the example of the Billy, and of the dog barking, are taken by many to be *spurious* 'causes' that arise from adopting a fragile view of events. However they only arise if the fragility applied is extreme—where *any* variation in timing or manner makes for a new event—and if this extreme fragility is constant across contexts. If the extent of fragility varies across contexts, however, then the 'spurious' causes may not arise:

> So if we wanted to make away with the stock examples of [pre-emption], what we would require is not a uniformly stringent standard of fragility, but rather a double standard—extremely stringent when we were trying to show that an effect really depends on its alleged [pre-empted] causes, but much

> more lenient when we were trying to agree with common sense judgements that an effect is not caused by just anything that slightly affects its timing and manner. It is not out of the question that there should be such a double standard. But if there is, an adequate theory of causation really ought to say how it works. To say how the double standard works may not be a hopeless project but for the present it is not so much unfinished as unbegun. [Lewis, 1986d, p.199]

The CCT view I argued for in Chapter 2 offers just such a double, or fluctuating, standard of fragility—fluctuating, that is, with context. For the CCT view to assist with cases of late pre-emption it would need to be the case that the context of pre-emption cases somehow invoked a fragile reading of the events in question. I think they do exactly that.

Once we realise that the appropriate counterpart relation for the effect event is ambiguous, Maxim 1 is invoked. This justifies the reading of the event in such a way as to render the causal claim that Suzy caused the window to break true. The second maxim stipulates taking the effect event to be no more fragile than is required so, on the assumption that the dog bark, or the gravitational pull that Billy exerts, are not depended upon until much more fragile readings of the event are invoked, simply interpreting the claim as fragile enough to make Suzy the cause does not commit the speaker to accepting that the dog bark is a cause within the same context.

However, at no point is the time difference between Suzy's rock and Billy's specified. For all the example tells us, the gap could be a second or a millionth of a second, so there will be late pre-emption cases that we can construct in which the difference between the two candidate events (s and b) is too small for any observer to plausibly detect. Maxim 3 would therefore seem to bar interpreting the speaker as if they can differentiate the window break s from the window break b.

Perhaps an observer could not plausibly detect the gap between Suzy's rock landing and Billy's but the problematic pre-emption cases being considered here are described, not observed. In the set up of the case we are told that Suzy's rock lands before Billy's and that his travels through the void. This is all that we require in order to know that there is a gap between their rock's landing, even if we do not know what the extent of that gap is. Say the gap is an unspecified x seconds, then given that the set up includes the information that Suzy's rock reached the window x seconds before Billy's, Maxim 3 permits that we interpret the causal claim that Suzy caused the window to break in line with this information. To see that we really are utilising this information, consider the same case re-described without this detail:

8. Billy and Suzy are throwing rocks at a window. A rock reaches the window, the window breaks.

Here the story is told without the chronological element concerning the two rocks and on this description there is no justification for selecting Suzy over Billy as a cause of the window breaking—the best that can be claimed is that 'a child' caused the window to break. So the intuition that the counterfactual analysis seemingly fails to

match (that Suzy is the cause) tracks the presence of a gap between Billy and Suzy. A contexualist account of causation has the resources to incorporate this important clue.

First, take a contrastivist account of causation which takes the contrast events $c*$ and $e*$ to vary with context. If such a view remained insensitive to the presence of the gap it may construct the following false contrastive claim:

9. Suzy's throwing the rock, rather than dropping it, caused the window to break rather than not break.

This is a false claim because had Suzy dropped the rock, the window would still have broken thanks to Billy. Presumably this interpretation of the context is what makes Schaffer believe that his contrastivism does not have the resources to resolve pre-emption problems [2005, p.358, note 35]. However, if we pay close attention to the clue given—that there was a gap between the rocks landing—then the following true contrastive contrastive claim is more in line with the information at hand:

10. Suzy's throwing the rock, rather than dropping it, caused the window to break as it did rather than break later.

This contrastive claim is true because had Suzy dropped the rock the window would have broken later courtesy of Billy. So, a suitably sensitive contextualist account can use the content of the set-up to charitably interpret the causal claim.

On the rival contextualist account of causal claims that I am advocating, the CCT view, the presence of this gap impacts on the truth conditions of the causal claim that Suzy caused the window to break as it changes the counterpart relation for the window breaking event to include only those breakings that happened within the period of the gap between Suzy and Billy, whatever that gap may turn out to be. The maxims introduced earlier justify this reading.

3.3.2 Disputes and Relevance

We can imagine Billy and Suzy's parents disputing who broke the window. Perhaps Billy's parents claim it was Suzy who broke the window and Suzy's parents claim that it was Billy. On the assumption that Billy's gravitational influence on the window was such that his throwing the rock made a minute difference to the timing or manner of the window breaking, there is some extreme level of fragility under which Billy's throw *is* a cause of the window breaking. So both pairs of parents speak truly according to the view I have presented. This looks like a problem as we would ordinarily distinguish Suzy's role from Billy's, especially when blame was being apportioned.

Of course Maxim 3 will rule out interpreting the requisite level of fragility for Suzy's parents' claim—they simply do not have access to the level of detailed evidence that they would require in order to detect Billy's influence on the window breaking. However if Suzy's parents are physicists, or even just causal-savvy philosophers, they can reason that Billy would have some such influence, much as I have been reasoning based simply on the descriptions and the rudiments of current physics. That means that Maxim 3 cannot block such a claim if the person in question has a certain level of education

or knowledge about the subject matter being discussed.[11] Perhaps they were being pedantic, and it is the right interpretation, or perhaps they did not mean such a fragile level after all.

Suppose that they are being pedantic: they are claiming that Billy's throw had some minor influence on the window breaking event. However in the context of apportioning blame, or working out how to avoid similar things happening in the future, citing the gravitational influence of Billy is simply irrelevant. If that was the standard that is being applied then Suzy, the dog who barked and even the parents several miles away had a similar influence. Being pedantic in this way flouts the Gricean maxim of relevance.

So, when Suzy's physicist parent seeks to blame Billy, we can either take them to be being pedantic, and therefore irrelevant, or take them to speak falsely. Either way, the claim should be dismissed. In ordinary cases we should not take the person to be so pedantic as to render what they are saying irrelevant to the discussion, it would be kinder to take them to be mistaken about the matters of fact. That is why even the physicist should not be taken to speak any more fragilely than the discourse requires, just as the first maxim says.

Whatever interpretive maxims apply in normal discourse will also apply if that discourse contains a causal element. Insofar as the task of giving a complete interpretive metric is incomplete for normal discourse, it will remain similarly incomplete for discourse with a causal element. However, the maxims that I have introduced show which specific element of the context is open for interpretation in causal statements and how we can most cooperatively interpret it. There will be problem cases for discourse with a causal element, just as with Suzy's physicist parent, but where the structure of such cases is to be found in non-causal discourse too, then such a case is not a problem for the causal theory on offer, but a problem instead for the pragmatic theory being employed. Pre-emption cases seem like a causal-specific issue and so a theory of causation must account for them, but pedants are everyone's problem.

3.4 Causes and Proportion

What I have said so far concerns the impact of varying the modality of the effect event, in particular that as you tend to interpret the effect event as strictly more fragile, you tend to render the causal claim true. The inverse is the case for the cause event: the more fragile the cause event, the more the claim will tend to be false. For example, 'slamming the door stopped the draft' does not seem quite the right thing to say, since any form of closing the door that wasn't a slam would have achieved the same end of stopping the draft. In fact, if you interpret the cause event in a more fragile way (holding the fragility/robustness of the effect fixed) the claim will likely be false, since the closest alterations of 'slamming the door' are very similar alterations which will doubtless bring about a very similar effect. Similarly, citing the scarlet colour of the patch that made Sophie the pigeon peck, when she was in fact trained to peck any shade of red patch, is peculiar to say the least, and false if you consider the cause event

[11] I thank Daniel Nolan for this example and for pressing the objection.

to be fragile since this renders the closest alternatives too close: the scarlet patch in the actual world is replaced by an almost-scarlet one in the closest worlds where the fragile cause does not occur.[12]

Conversely, describing the cause events in more general terms renders them more robust and tends to make the claim true: moving the door caused the draft to stop, the coloured patch caused Sophie to peck. Of course, not all movements of the door, and not all colours of patch, would bring about the respective effects, but absent *any* movement of the door the draft is not excluded, and a colourless patch would not make Sophie peck. Further, if we know that closing the door, rather than any movement at all, would exclude the draft then it would be odd to be less than specific about it. Similarly if we know that the patch needs to be red for Sophie to peck, then it would be odd not to specify that.

The maxims I offered earlier were derived from general principles of interpreting what someone has said, but were specific in relation to the effect event. A different specification of the first two maxims is required for the cause:

Maxim 4 *Take the cause event to be as robust as is required to render the causal claim true.*

Maxim 5 *Take the cause event to be no more robust than is required to render the causal claim true.*

These maxims justify the interpretation of the cause event in the examples respectively as essentially a closing, and only accidentally a slamming, and as essentially red and only accidentally scarlet.

This double-standard for cause and effect seems problematic. If causation is transitive, as I will assume it is for the moment,[13] any event which is an effect can also be a cause. If we have conflicting standards for the treatment of an event when it is a cause and when it is an effect, what are we to say when it is both? Perhaps there is no single context in which a cause is both, but if there is it may well be a restriction in that context that the event meet all of the maxims (1–5) simultaneously. Is this possible? I think it is. Sophie the pigeon was trained to peck all and only red things, so specifying that the patch was coloured flouts Maxim 1 and specifying that the patch was scarlet flouts Maxim 4. However specifying that the patch was red respects each of the maxims: it is fragile enough, without being too fragile, and robust enough without being too robust. We do not ordinarily require that a causal claim meets this exacting standard in order to be acceptable but we are supposing that there is a context in which the effect is both a cause and an effect so it may just be a feature of such contexts that they have a far stricter standard of causal ascription than ordinary contexts. I will argue that this is exactly the case when I come to discuss the transitivity of causation in Chapter 7.

Before moving on, let me point out an interesting and useful parallel with a proposal from Stephen Yablo [1992]. For Yablo the specification of the cause should neither give

[12]This is a particularly important issue quite apart from considerations of pre-emption. I will return to it in Chapter 4.

[13]I discuss this more fully in Chapter 7

too much information, nor too little, with respect to the effect.[14] Specifying that Sophie was presented with a scarlet patch gives too much information, specifying that it was a coloured patch gives too little, but specifying that it was a red patch gives just enough information. This requirement is known as Yablo's *proportionality* constraint. For the moment I simply want to point out this parallel with the proposal I offer and point out a couple of important differences.

Firstly, if we take the proportionality constraint as a constraint on which causal claims are true then it will provide an extremely high bar for correct causal talk: it will be false to say that the scarlet patch caused Sophie to peck. However, if it is a constraint on which causal claims are *optimal* then it is a pragmatic constraint which is a rival to, and largely concordant with, the maxims I have advocated.

Secondly, Yablo's constraint only makes reference to the cause and not the effect. Perhaps the constraint could be extended so it applied to either side of the causal relation so long as proportionality is reached in the end. It is not clear, though, how this would help in cases of late pre-emption. In those cases it was not just that the causal claims were out of proportion, it was that the context had justified a more precise reading of the effect event in particular. Proportionality could have been achieved instead by making the cause (Suzy's throw) less precise (a child's throw), but this would have been to avoid the pre-emption problem that the counterfactual analysis faced rather than resolve it. On the view I have presented, the set up of the case—in which we know that there were two children throwing—rules out reading the cause more robustly (a child's throw) but justifies reading the effect as more fragile (a window breaking at that time and in that way). Yablo's constraint, in its current form, lacks this important sensitivity to context.

3.5 Conclusion

In this chapter I have argued that Lewis's 'dare to be a doctor' riposte against the fragilicist is misjudged as it conflates two different readings of the causal claims being made. Lewis's related complaint about spurious causes (such as Billy and the gravitational pull of Jupiter each being causes of the window breaking) requires us to believe that someone who endorses a fragile view of events in the pre-emption context is committed to a similarly fragile view in all contexts. This simply does not follow. It does, however, establish an explanatory burden for any view that wishes to hold that effect events are fragile in pre-emption cases but not in others.

I have proposed a view of the pragmatics of causal talk that provides just such an explanation. Taking some relatively basic pragmatic maxims from Grice and Lewis, I have argued that certain causal-specific maxims can be derived (given a counterfactual theory of causation). Applying those maxims in the context of pre-emption justifies a fragile reading of the effect events without entailing a commitment to fragile events in all contexts.

[14] This is a crude gloss of the view to be supplemented in Chapter 7 when I discuss transitivity.

4

On the Non-occurrence of Events

In the previous chapter I argued that a counterfactual analysis of causation could give the right results in the apparently problematic cases of late pre-emption, so long as a certain contextualised reading of the effect event was deployed. This solution concerned the fragility of the events in question and required that we be sensitive to the impact of context upon the modality of the effect event in question. Shifting the modality of the event shifted the conditions under which the event was taken to occur or not occur.

Of course many effects will go on to become causes and I argued in the last chapter that causes and effects required inverted pragmatic maxims, though each is still governed by the same super-maxim: assume that the conversational contribution is true. In this chapter I further explore the issue of event fragility and reach an interesting and surprising conclusion: there is an asymmetry in the standard of non-occurrence applied to cause events and effect events when reasoning counterfactually about them.

4.1 Recap

To begin, it is worth recapitulating the story so far. Lewis's original [1973] counterfactual analysis of causation states that one event is a cause of another iff there exists a chain of counterfactual dependence between them. One event e counterfactually depends on another c iff the following counterfactual conditional (or causal conditional) is true:

$$\neg Oc \; \Box\!\!\rightarrow \neg Oe$$

In Chapter 2 I argued that the truth of this conditional could shift with context and expression, suggesting that there is a variable in the semantics that is not represented in Lewis's test. I endorsed a counterpart-theoretic view of events which would explain this variation and introduced new notation to help track this variable. Events c and e

under counterpart relation n would be denoted by c_n and e_n respectively. This yielded the following revision of Lewis's test for causal dependence between events:

$\neg Oc_n \; \Box\!\!\rightarrow \neg Oe_n$ (where n is provided by the context)

The revised analysis of causation that this then forms (together with the chaining requirement) I named CCT for Counterpart-theoretic Counterfactual Theory of causation.

This approach helped me disambiguate two different claims that we could endorse in cases of late pre-emption. In the Billy and Suzy example, it allowed me to distinguish the window breaking *exactly as it did* and the window breaking *in some way or other*. I argued that we could resolve the late pre-emption problems by tending towards a more fragile (more restricted) reading of the effect event's counterpart relation (I also argued that such a reading was justified under a Gricean/Lewisian pragmatic framework).

So-called 'fragile' events are taken to have an abnormally rich essence, but quite what a normal essence would be is far from obvious.[1] In any case events under a relatively strict counterpart relation will occur in relatively few worlds and will fail to occur in relatively many. In other words, the negation of the proposition *e occurs* is true in a larger range of worlds when e is taken to be fragile than when it is not. This means that when e is considered fragile the consequent of our causal counterfactual is weaker than when e is considered robust, and that in turn makes the conditional easier to satisfy. So, tending to interpret the effect event as fragile in turn tends to yield a true result in our CCT test. Thus the CCT test, combined with a fragile view of events, is very permissive about when an event is to be considered a 'cause'.

4.2 Fragile Causes and Excision

Fragile effects are one type of problem, but fragile causes are a different type. Where fragile effects weaken the consequent of the causal counterfactual, and yield seemingly too many causes, fragile causes weaken the antecedent and yield too few. On a fragile reading the version of Suzy's throw that takes place in a nearby world, in which Suzy's hair (or even just one of her hairs) is a slightly different shade, is a different event. This nearby world then satisfies the antecedent of the causal counterfactual $\neg Oc$. However even on the most fragile standard the effect event is unchanged. Had Suzy been slightly larger, or shouted as she threw, such alterations could, in principle, be detectable in the effect event by the difference in gravitational force or resonance. Assuming for the sake of argument that the change in shade is not a difference that requires a gravitational shift,[2] the shade of Suzy's hair in no way influences the window's breaking. Here we have a very close $\neg Oc$-world in which Oe is true.

[1] An essentialist may be able to claim that there is just such a 'normal' essence that applies but I am operating under an anti-essentialist assumption by adopting a counterpart-theory of events. Nevertheless, I will argue in Chapter 5 that there is a nominalist-friendly route to identifying a default or canonical counterpart relation.

[2] It is not obvious whether or not this is plausible within our physics since it requires that the shade be altered (presumably by altering the physical make-up) without an attendant shift in gravitational influence. The ubiquity of gravitational influence in our physics muddies certain examples and so I

Is it one of the closest worlds? Perhaps not as I have described it, but there will be some alteration of c which concerns shade and which is sufficiently minimal so as to occur in one of the closest $\neg Oc$-worlds. That is all that is required to show that on the most fragile reading, the claim c *caused* e is false. It only takes one $\neg Oc$-world in the closest sphere to be an Oe-world for the counterfactual analysis to deny that there is a causal connection between one event and another.

One might think: so much the worse for the fragile view. However the problem is not restricted to fragile cases. Once we see how the problem is generated we can reproduce the issue at less fragile levels. Here is the formula: identify some essential aspect of the cause event which, if varied, would have no impact on the effect event. For example, assume that Derek's throwing the heavy red ball caused the bottle to break. Regardless of how fragile the treatment you apply to the bottle breaking is, having Derek throw a heavy green ball instead would make no difference to the effect event—again, I am stipulating that we keep gravitational complications in abeyance by altering the shade. Throwing a heavy red ball is not the same as throwing a heavy green ball and so the nearby world in which Derek throws the green ball is a $\neg Oc$-world. Is it one of closest $\neg Oc$-worlds? Perhaps not as I have described it, but there will be some alteration of c which concerns the shade, and which is sufficiently minimal, so as to occur in one of the closest $\neg Oc$-worlds.

The problem in these cases is that Lewis's counterfactual test fails if any of the essential aspects of the cause are irrelevant to the effect. The requirement for a successful causal claim on this test is that every aspect of the cause event impacts in some way on the effect. Yet when we ask if one thing caused another, we are not asking if every aspect of the first impacted the second, but rather if any aspect of the cause had an impact on the effect. As it stands, the Lewisian test is inadequate for that purpose.

4.2.1 Clean Excision

Noting this problem, Lewis (1986d, p210-211, 2004a, p.90) points out that it stems from ambiguity in what we mean by c's non-occurrence. If we accept that any alteration, any almost-c counts as a non-occurrence of c, then the problem discussed above emerges. Lewis points out that such problems are avoided if we imagine that c is "cleanly excised from history, leaving behind no fragment or approximation of itself". The idea here seems to be that when considering the non-occurrence of an event, we ought to imagine that nothing remotely like it occurs. The proposal clearly lacks detail but I will run with it for the moment and return to the detail in §4.2.3.

This *clean excision* policy leaves the original analysis intact but amends the interpretation of $\neg Oc$. Applying this in the problem cases of Suzy and Derek we see that it does resolve the problem. Ignoring Billy for the moment, had Suzy not thrown the rock *at all*, the window wouldn't have broken, and if Derek had not thrown the heavy red ball *at all*, the bottle would not have broken.[3] These are the right results.

use the device of altering the colour to side-step this contingent complication. If I were offering a theory restricted to the physically possible worlds this would be a potential problem but since I am not, I only need it to be conceptually, rather than physically, plausible.

[3]The emphasis of 'at all' is to signify that no close alternative is substituted.

They are the right results so long as Billy remains ignored, but that would be to ignore the problem of pre-emption that motivated the discussion of fragility in the first place. Recall that our solution to the pre-emption problem was to specify the effect event under such fragile standards as would allow a counterfactual test to detect the difference between Billy and Suzy's throws. This fragile approach distinguishes between the window breaking in any way at all and the window breaking exactly as it did. Combining this with the clean excision policy from above gives us: had Suzy not thrown the rock at all, the window wouldn't have broken exactly as it did. Again, this is the right result but note that we have applied a double standard—a clean-excision policy to the cause and a nearest-alternative policy to the effect. This is not a mere quirk. If we apply either standard universally, we get false claims in straightforward examples:

Clean Excision Had Suzy not thrown the rock at all, the window wouldn't have broken at all.

Closest Alternative Had Suzy not thrown the rock exactly as she did, the window would not have broken exactly as it did.

Clean Excision is false because Billy would have brought about a close alteration of the window breaking. In fact it is false even if the window breaks a century hence, and so the problem is not unique to the standard pre-emption cases.

Close Alteration is also false: had Suzy not thrown the rock exactly as she did, the window would not have broken exactly as it did. This is false because, as shown above, changing the hue of the rock, or the colour of Suzy's hair counts as a nearby alteration (making $\neg Oc$ true) but makes no difference to the effect event (leaving Oe true). Again, this is not restricted to pre-emption cases.

From this I conclude that the solution to the pre-emption problem requires a mixed, or double, standard of non-occurrence.

4.2.2 Two Standards of Non-occurrence

Such a double standard may seem theoretically untidy but not only does it offer a solution to the pre-emption problem, I believe there is also independent motivation for this asymmetrical treatment of cause and effect which justifies adopting the double-standard.

The two standards under consideration can apply in four combinations. Both cause and effect can be held to clean excision standards (CE-CE); both cause and effect can be held to closest-alternative standards (CA-CA); cause can be held to a clean excision standard and effect to a closest alternative standard (CE-CA); or vice versa (CA-CE).

Where either cause or effect is held to a clean excision standard, all of the aspects of that event must be absent in its negation (i.e for its non-occurrence). Where either one is held to a closest alternatives standard, the negation of the event-occurrence is satisfied when any single aspect of that event is absent. Thus we can phrase our four alternative analyses as follows.

c is a cause of e if and only if:

CE-CE altering all aspects of c alters all aspects of e

CA-CA altering any aspect of c alters at least one aspect of e

CA-CE altering any aspect of c alters all aspects of e

CE-CA altering all aspects of c alters at least some aspect of e

It will be useful to test these alternatives in light of the example of Derek who threw the heavy red ball. Clearly, Derek's throw caused the bottle to smash. A causal analysis must agree with this claim to be acceptable.

CE-CE will be too strong a requirement as even the complete excision of Derek's throw would leave some aspect of the bottle smash unaltered: the transparency of the glass, the presence of a particular molecule, or the pattern on the base. CE-CE gets the test case wrong.

CA-CA is also too strong a claim. Altering the hue of the ball does not alter the bottle breaking event in any way. CA-CA gets the test case wrong.

CA-CE is an even stronger claim than CA-CA. If altering the hue doesn't alter any aspect of the bottle break, then *a fortiori* it cannot alter all aspects of the bottle break. CA-CE gets the test case wrong.

CE-CA is the weakest claim. CE-CA is satisfied if the clean excision of the cause, i.e. Derek's throw, makes any difference to the effect, i.e. the bottle breaking. CE-CA is the only standard that gets the test case right.

Note that this is a simple test of the available combinations once we realise that there is an ambiguity in the notion of a non-occurrence of an event. The test case used is not a pre-emption case and so the results of the test case are independent of my argument concerning the required treatment in pre-emption cases. However, this independent assessment points to the same conclusion: the Lewis analysis must be amended to reflect a double standard in the treatment of non-occurring causes and effects.

One reason to think that this double standard should not apply comes from reflecting on the chaining nature of causation. If c causes d and d causes e, then d is both a cause and an effect.[4] If one event can be both a cause and an effect, then there should be no asymmetry between the ontological status of causes and effects. Note, though, that I am not advocating an ontological difference between cause and effect, nor even a difference in counterpart relation, I am advocating an asymmetry in our analysis of causation–that is, we apply one standard of non-occurrence to d when it is in the cause role, and another when it is in the effect role. I will return to a deeper worry about transitivity briefly in §4.4 and more fully in Chapter 7.

4.2.3 Refining Clean Excision

The notion of a clean excision is doing a lot of work in the preceding discussion but it is very much underspecified by Lewis. In this section I propose a definition for each of the alternative standards of non-occurrence.

[4] I owe this point to Schaffer [2005].

The first thing to note about the CE and CA standards is that they are defined in terms of event *aspects*. I am assuming a coarse-grained conception of events and those events have properties which are synonymous with the aspects of the event. I am also assuming a counterpart theory for events. According to this view an event described as 'the throwing of the red ball' could easily be the very same event as described by 'the throwing of the heavy ball'—it will be the same event if it occupies all and only the same region of a world—but the first description is likely to invoke a counterpart relation in which each counterpart involves a red ball (regardless of weight) but the second is likely to invoke a counterpart relation in which some counterparts are not red but in which all are heavy. The aspects of an event that are shared by all of its counterparts are *essential* and the rest are *accidental*. Importantly, it is a context-sensitive matter which counterpart relations obtain and, as argued in Chapter 2, this context sensitivity tracks the context sensitivity of our causal ascriptions.

This context sensitivity would be lost if our causal test were not also sensitive to which aspects of an event were essential and which were accidental. If *every* aspect of the event were on a par, then the clean excision standard of non-occurrence of the event would be the same regardless of how fragile or robust the event was. It would mean that there was no difference between cleanly excising a fragile event and a robust one which would undermine what I have said about pre-emption cases where the robustness/fragility of an event makes a significant difference. Under a counterpart theory of events, however, we can distinguish the different aspects of the event under consideration. So, I propose the following definitions:

Fully-occurs An event c occupying region R of world w is taken to fully-occur (Oc) at world w_n iff there is some region of w_n which corresponds to R in w and in which all of c's essential aspects are instantiated.

Partially-occurs An event c occupying region R of world w is taken to partially-occur (Pc) at world w_n iff there is some region of w_n which corresponds to R in w and in which at least one of c's essential aspects are instantiated.

This allows us to define our two standards of non-occurrence, CE non-occurrence and CA non-occurrence, as follows:

CE non-occurrence Actual event c CE non-occurs at world w iff $\neg Pc$ is true at w.

CA non-occurrence Actual event c CA non-occurs at world w iff $\neg Oc$ is true at w.

In other words, a close alteration of c exists in worlds where even one essential feature of c is present in the appropriate region but c is cleanly excised in worlds where no essential feature of c is present in the appropriate region.[5]

[5]I note here an issue with specifying the corresponding region in another world. Perhaps regions can have counterparts as picked out on the basis of some extrinsic spatiotemporal features. This would fit with Lewis's belief that some minimal extrinsicality was required in specifying events—spatiotemporal location and the laws were both bound up in the identity condition for events, according to his [1986d, p.264].

It is interesting to note that in Lewis's original specification of the counterfactual analysis of causation he used the propositions Oc and Oe but in subsequent discussions it has become commonplace to drop the O for brevity. However, given the ambiguity regarding the standards of non-occurrence of an event perhaps this short-cut was not harmless.

4.3 Retrospective

In retrospect it is easy to see how this ambiguity remained so well hidden. Firstly, the O representing 'occurs' is typically dropped in discussions of counterfactual analyses so the pivotal variable (as I have identified it) is almost always glossed over in the relevant literature. Second, English only has a limited range of locutions for specifying the key idea: the non-occurrence of the event, the event does/did not occur. These locutions do not track the distinction between the different standards of non-occurrence in question. Third, the two standards very frequently converge. This last point is worth demonstrating in more detail: when the cause event is specified in such a way as to trigger a counterpart relation with a single essential aspect, it makes no difference whether you consider the $\neg Oc$ scenario to require a counterpart without *any* of the essential aspects of c (CE standard), or simply a counterpart that lacks just *one* essential aspect of c (CA standard). Since there is just one essential aspect the absence of one aspect or all amounts to the same thing and the two standards of non-occurrence converge and give equivalent results when plugged into a counterfactual conditional. More generally, whenever the closest $\neg Oc$-worlds and $\neg Pc$-worlds are the same, the CA and CE standards of non-occurrence will give equivalent truth conditions for the causal counterfactual that takes c to be the putative cause.

Even when the closest $\neg Oc$-worlds and $\neg Pc$-worlds are different the causal counterfactual can still yield the same results. For example, suppose that Suzy's throwing the rock causes the window to break and further suppose for simplicity that the cause event has just three essential features: it involves Suzy, a rock and at throw. The clean excision of this event requires that there be no Suzy, no rock and no throw. When c is cleanly excised, the window doesn't break and so the causal counterfactual comes out true on the CE standard of non-occurrence of the cause ($\neg Pc$). On the CA standard of non-occurrence, $\neg Oc$ fills the antecedent of the counterfactual. The worlds where c does not fully occur include those where Suzy drops the rock, where someone else throws the rock and those where Suzy throws some other object. Of those $\neg Oc$-worlds, which is the closest? In this example it seems likely that the closest $\neg Oc$-worlds are those where Suzy drops the rock rather than worlds in which she morphs into a different person or where the rock is suddenly supplanted by some other object. If the closest $\neg Oc$-worlds (where Suzy drops) are all worlds where the window doesn't break, then the causal counterfactual will be true in that case. Of course I have manipulated a toy example here but the point to make is that there will be some examples with this form and in such examples whether the CE and CA standards of non-occurrence are applied to the cause will make no difference to the truth of the causal conditional. This further explains why the distinction between the standards remained hidden.

If the two standards can and do agree, then perhaps it was an artefact of the example chosen above that the CE-CE, CA-CA and CA-CE standards were rejected in favour of the CE-CA standard. However, whilst these standards will all converge when there is a single essential aspect of both the cause and the effect, they will not always converge when the events in question have multiple essential aspects. This is all that is required to justify distinguishing the standards. To illustrate: suppose that we have a cause event with three essential features (P, Q and R) and an effect event with three essential features (S, T and U). For the purposes of notation, I will define c_P as the occurrence of c, or a counterpart of c, with feature P and c_Q as the occurrence of c, or a counterpart of c, with feature Q, and so on. Now we can state the truth conditions for each of our four standards of non-occurrence in this example:

CE-CE will be true if and only if $\neg(c_P \vee c_Q \vee c_R) \boxdot\!\!\rightarrow \neg(e_S \vee e_T \vee e_U)$ is true, that is if, in all of the closest possible worlds, the total absence of any member of the set $\{c_P, c_Q, c_R\}$ correlates with the total absence of any member of the set $\{e_S, e_T, e_U\}$.

CA-CA will be true if and only if $\neg(c_P \wedge c_Q \wedge c_R) \boxdot\!\!\rightarrow \neg(e_S \wedge e_T \wedge e_U)$ is true, that is in all of the closest possible worlds where any member of the set $\{c_P, c_Q, c_R\}$ are absent, at least one of the set of set $\{c_S, c_T, c_U\}$ will be absent.

CA-CE will be true if and only if $\neg(c_P \wedge c_Q \wedge c_R) \boxdot\!\!\rightarrow \neg(e_S \vee e_T \vee e_U)$ is true, that is in all of the closest possible worlds where any member of the set $\{c_P, c_Q, c_R\}$ are absent, no member of the set $\{e_S, e_T, e_U\}$ will be present.

CE-CA will be true if and only if $\neg(c_P \vee c_Q \vee c_R) \boxdot\!\!\rightarrow \neg(e_S \wedge e_T \wedge e_U)$ is true, that is if, in all of the closest possible worlds, the total absence of any member of the set $\{c_P, c_Q, c_R\}$ correlates with absence of any single member of the set $\{e_S, e_T, e_U\}$.

In a more concrete example, assume that a player shoots low (P) and hard (Q) wearing purple boots (R) and that the keeper dives low (S), quickly (T) wearing yellow gloves (U). Assume that had the player shot otherwise (high, soft, off target) the keeper would have dived otherwise (high, slowly, not at all).

The CE:CA standard gets this case right: By the CE-CA standard the dive is caused by the shot because in the total absence of any of the essential features of the shot (P, Q or R) at least some aspects of the dive are altered it is no longer low and quick. Thus the total excision of the shot alters the dive in some respect and this makes the shot a cause of the dive.

All other standards get the case wrong: By the CE-CE standard the dive is not caused by the shot because the gloves remain yellow (and so e_T still occurs) even if the shot is cleanly excised. By the CA-CA standard the shot does not cause the dive because even in worlds where the boots are not purple (and so c_R is absent) all essential aspects of the dive (S, T, or U) remain unaltered. By the CA-CE standard the dive is not caused by the shot because for every alteration of the shot (c_P, c_Q, or c_R), some alteration of the dive (e_T at least) occurs.

I think that this example shows how the cases come apart with multiple essential features being attributed to the cause and effect and justifies (i) distinguishing the two standards, and (ii) applying those standards asymmetrically (CE-CA) in the causal counterfactual. I now turn to the implications of this result.

4.4 Revising the Counterfactual Analysis

If what I have said about the requirement of a double standard of non-occurrence is correct, then I will need to propose an alteration to Lewis's original counterfactual analysis. More specifically I will need to alter the conditional that he claimed established causal dependence. Lewis said that c is a cause of e iff (i) c and e are distinct and (ii) there exists a chain of causal dependence between c and e where causal dependence was established by the truth of the following counterfactual conditional:

Lewis $\neg Oc \boxdot\!\!\rightarrow \neg Oe$

I have argued (Chapter 2) that we needed to amend this test to incorporate the counterpart variable for events. Using subscripts to indicate different counterpart relations, this yielded the following modified test for causal dependence (which I called CCT). e causally depends on c relative to counterpart relation x iff the following counterfactual conditional is true:

CCT $\neg Oc_x \boxdot\!\!\rightarrow \neg Oe_x$

In ordinary causal discourse, the value x is set to a specific value, n, which is determined in part by the context of utterance and the mode of representation of c and e.

However, as we have seen O can be read either as *partially* occurs or as *fully* occurs depending on the standard of non-occurrence that applies. The foregoing argument is intended to show that applying either reading of O to *both* c and e will not give intuition-matching results but that applying the first to c and the second to e will. Hence I have proposed an asymmetrical standard for the non-occurrence of the cause and the effect respectively. In place of my revised CCT test for causal dependence from Chapter 2 I offer the following amendment: e causally depends on c relative to counterpart relation x iff the following counterfactual conditional is true:

ACCT $\neg Pc_x \boxdot\!\!\rightarrow \neg Oe_x$

Here, I supplanted the first O for occurs with a P for part-occurs as defined above. This amounts to applying a clean excision standard to one side of the conditional and a closest-alternative standard to the other. The CCT test for causal dependence is now supplanted by what I will call an **ACCT** test, adding the A for Asymmetrical in reference to the asymmetric standard of non-occurrence being applied to the cause and the effect respectively.

When the context in question is the ordinary context of causal talk, I will refer not just to the ACCT test for causal dependence (which allows for *any* counterpart relation to fill x) but to the **ACCT Contextual** test which takes variable x to be set to value n, the counterpart relation invoked by the ordinary context of our causal discourse. This will prove a useful distinction in Chapter 5.

Causation is still analysed as chains of causal dependence but my analysis of causal dependence now reflects the findings of the previous chapters: that counterpart variation must be represented and that there is an asymmetric standard of non-occurrence that applies to the cause and the effect in the relevant counterfactual conditional.

This is a novel way of expressing the asymmetry between cause and effect but positing just such an asymmetry is familiar from Paul [1998a]. Paul advocated taking the effect event to be sensitive to time in such a way as to render the late pre-empting Suzy to be considered a cause, but not Billy. No such sensitivity is applied to the cause however. Lewis [2004a] extended Paul's idea to include both the timing or the manner of the effect to be salient to whether it had counterfactually depended upon the cause. This newer theory from Lewis advocates using a 'tailor-made' proposition about whether, when and how e occurred as the consequent in the counterfactual conditional. When the counterfactual is true then that means that *whether*, *when* and *how* e occurred depends on *whether* c occurred. This is very similar to the notion I have adopted above—in my version we hypothesise c's not-occurring *at all* and look to see whether e happened at all, at a different time or differently.

Lewis's modification doesn't stop there however. Rather than leave an asymmetric test that treats cause and effect differently, Lewis argues that we ought to look at a range of alterations of c—alterations where c didn't occur at all, where c occurred at a different time and where c occurred differently—and ask whether e happened at all, at a different time or differently in each case. This amounts to mapping a range of non-actual alterations of c onto a range of non-actual alterations of e. If it is true that at least some *not-too-distant* alterations of c correspond to different e alterations, then c can be said to *influence* e. In this revised account, causation does not require chains of causal *dependence* as it was in the original analysis, but rather chains of causal *influence* understood in this revised way. The later analysis is weaker as every case of dependence is a case of influence, but not vice versa.

I have two complaints about this revised theory from Lewis. First, it is unclear about what the range of alterations under consideration are. Lewis obviously wishes to rule out alterations to c which create a nuclear explosion or a black hole but, unlike his account of *the closest possible world* in which $\neg c$ is true, which he defends at length in [1979], this idea of not-too-distant is vague and underspecified: it leaves intuition doing all the running. However, even if a good account is given at to what the salient alterations are, Lewis seems to have changed the subject from *actual* causation to *actual*-or-*possible* causation. No longer are we focussed on the occurrence of c in the actual world and what it achieved, we are now concerned with what some imagined alteration of c *would have* achieved. I didn't cause a riot, but if I had acted differently I would have. I take this distinction between actual and possible causation to be important and Lewis's later theory collapses that distinction. I will return to this issue of actual and possible causation when I discuss absence causation in Chapter 6 and I will make similar complaints against the causal modelling project when I come to look at it in Chapter 9.

Before moving on, I will note a problem with my asymmetrical approach. If the cause and the effect are subjected to different standards then an event which is the effect at one link in the chain could promptly become a cause in the next link. So, when the player shoots and the goalkeeper dives, then we might say that the shot caused the dive since the clean excision of the shot would result in the keeper's not diving. But we might also say that the keeper's presence in the goals at that point caused someone to think about a particular shade of yellow (from the gloves) since had the keeper been

cleanly excised, the thought would not have crossed their mind. If we further suppose that the thought about the yellow would have occurred whether the keeper dived or not (the gloves being visible either way), then it should be clear that the player's shot did not cause the thought about the shade of yellow. Yet the dive is caused by the shot and the diving keeper causes the thought of yellow to cross the spectator's mind so, on the assumption that causation is transitive, the shot *does* cause the thought of yellow. Something has gone wrong.

Interestingly, Lewis discusses this problem in relation to his revised theory which, whilst symmetric, remains open to this sort of counter-example [2004a, p.93-96]. Lewis, by insisting on the transitivity of causation, bites this bullet. For my part, I note the issue now and leave it to one side until I return to the substantial topic of transitivity in Chapter 7.

4.5 Conclusion

I have argued here that pre-emption problems help to reveal a more general problem with Lewis's original analysis concerning an ambiguity in the conditions for the non-occurrence of an event. I have argued that we have independent motivation, quite apart from pre-emption cases, to distinguish two senses of non-occurrence and to apply them asymmetrically in our causal analysis. Lewis spotted this issue too but rather than clarify the notion of non-occurrence, he instead liberalised his theory in a way that collapsed the important distinction between actual and possible causation. That is a distinction I am not willing to collapse.

The ACCT view that I have developed over the last three chapters can account for contextual variation and pre-emption cases within a binary account of our causal talk. This relates to aim (I) of the thesis—to give an account of our everyday causal talk—and in the next chapter I will relate this discussion to my other aims: (II) to give an account of the mind-independent, objective standard for causal connectedness between events; and (III) to explain the relation of (I) and (II).

5

The Privileged Context

In the preceding chapters I have argued that the truth of certain causal assertions varies with context. I have also argued that such variation can be traced to shifting counterpart relations that apply to the events referred to in those assertions, and I have offered an outline pragmatics for interpreting this element of the semantics across different contexts. In short, I have argued for a contextualist reading of causal claims in our everyday talk in line with aim (I) of this thesis.

Such a contextualist approach might appear to be in tension with aim (II) of the thesis: to give an account of the mind-independent, objective standard for causal connectedness between events. According to Menzies, the idea that causation is just such a 'natural' relation is the central platitude of our causal concept (albeit one that Menzies does not subscribe to [2009]) and it is reasonable to think that this naturalism is incompatible with the mind-dependent, context-variant account of causal talk that I have been arguing for. If causal facts have objective truth conditions, how then can assertions about causation be true in one context and false in another?

In this chapter I aim to give an account of this natural relation in line with aim (II) and then go on to show that aims (I) and (II) are compatible. I will argue in favour of treating one particular context as *privileged* and I will argue that doing so allows us to hold both naturalist (in Strawson's sense of a mind-independent, objective relation [1992, p.109]) and contextualist views about causation and our causal assertions. I will show a deep connection between the *natural* causal truths that relate to aim (II) and the highly contextualised truths of our causal talk that relate to aim (I). This will allow me to offer a further revised analysis of causation and a closely related account of the truth conditions of our causal talk.

5.1 Tension

There is no direct conflict in advocating shifting truth values for our causal assertions whilst holding that the causal facts are fixed—the first posit concerns causal assertions and the second causal facts. If the subject matter is different then there is no direct contradiction.

However, it is obvious that causal claims and causal facts are related and it is incumbent upon the compatibilist to show how this relation avoids the seeming contradiction of contextualist and naturalist positions. For example, if causal claims were simply propositions whose truth supervened upon the causal facts, then there could be no change in the truth-value of a causal claim without an attendant change in the causal facts. Such a standard reading would give rise to a contradiction as it would rule out it being the case that causal truths vary with context whilst causal facts do not.

There may be no contradiction, though, if the causal claims conceal an implicit variable that tracks context and the causal facts do not. If a causal claim is not semantically complete without reference to a context then the variation across contexts of a single causal assertion, i.e. 'throwing the ball smashed the window', is not really the variation of a single causal claim at all but rather the same sentence expressing different propositions in different contexts. The same words can be used in one context where the ball in question is heavy and in a second context when there is a different ball which is light. We should expect the truth values of such claims to vary when the event being referred to varies.

Yet, those causal claims which are standardly taken to vary with context (i.e. those cases discussed in Chapter 2) do not vary *which* event is being referred to, but rather *how* that event is being referred to. Suppose that these two sentences are uttered in reference to a single serve by McEnroe:

> McEnroe's tension caused him to serve awkwardly.
>
> McEnroe's tension caused him to serve.

The first seems right but the second does not and yet the same actual cause (tension) and the same actual effect (serve) are being discussed. Contrastive approaches to contextual variation cases argue that there is indeed a missing variable in causal claims that tracks context, but that the variable does not shift which events are being picked out. On such views the actual-world relation of events remains the same across contexts but an implied contrast is generated by the causal context. So, when we talk of an awkward serve in a causal context, this implies a contrast scenario in which a graceful serve is executed. But when we discuss the serve without making reference to its awkwardness then this implies a contrast scenario in which no serve takes place at all, graceful or otherwise. Thus, the suppressed variable in causal claims is the implied contrast case, and it is the contrast case that varies with context.

This does not yet resolve the contextualist-naturalist tension. The contextualism on offer—contrastivism—does identify a suppressed component in causal claims, but that does not establish that the truth-value of a causal claim can vary independently

of the causal facts. In the McEnroe examples, the contrastivist holds that both claims concern the same actual-world events but bear different truth values. The naturalist will hold that there is a fact of the matter about the actual-world causal relation and so the two claims that pick out the same actual-world events can be expected to correspond to a single causal fact concerning those events. So, the tension remains.

Or at least it would remain if the proposed contrast cases were merely benign context variables, floating free of the causal facts. However, that is not the standard contrastivist position. Contrastivism does not simply hold that there are implied contrast cases in our causal talk, but rather that the structure of causal facts is itself contrastive. In Schaffer's version,[1] this amounts to the claim that causation is not binary as we might have expected, i.e. c causes e, but rather that it is quaternary, c rather-than c^* causes e rather-than e^*. This way a causal claim can concern the same actual-world c and e, as in the McEnroe case, but still vary with respect to c^* and e^*. The contrastivist position is that the implied contrasts vary with context and alter the causal fact being expressed (and therefore the truth value of the sentence stating that causal fact). Once the contrast is made explicit, and contextual variation removed, we have the objective standard of causal facts that naturalism required. The contrastivist has accounted for the context variation but also provided an objective standard of assessment for causal truth. The tension has been resolved.

Two issues fall out of this version of the contrastivist position. First, it requires that we jettison the assumption that causation is a binary relation. Second, it requires that causation not be a relation between actual events, but between actual events and possible events. I will not take issue with the content of these commitments here, but I will point out that they are revisionary. If the same theoretical gains can be made without requiring such revisions, then that will be preferable and I intend to show that they can.

5.2 Counterpart Variation

In this thesis, I present an alternative view which can endorse contextualism without revising our assumption of binarity (aim (I)), and which takes causation to be a natural relation between actual events (aim (II)).

Chapter 2 has already demonstrated the contextualist component. A reminder: causation is to be analysed as chains of counterfactual dependence linking distinct events. Causal claims include event expressions which invoke context-variable counterpart relations for the actual world events being picked out. As such, our event expressions, rather than implicit contrast cases, carry the context variable component of our causal claims. Thus, contextualism is endorsed whilst causation remains a binary relation.

[1] Hitchcock [1996] offers a significantly different form of contrastivism where only the effect triggers a contrast. Schaffer [2005] takes both the cause and the effect components in a causal claim to trigger contrasts, as do Menzies & List [2010], but unlike List and Menzies Schaffer considers the contrasts to be particular contrasts c^* rather than sets of contrasts C^*. It will serve me best to discuss Schaffer's version here since it is neatly and explicitly laid out in his Schaffer [2005].

In this section I argue for the naturalist component by showing that there is a 'natural' or canonical counterpart relation that applies to events mind-independently and where the context is undefined. This counterpart relation underpins the natural causal relation between actual world events. I will further argue that this natural relation has a particular logical priority: for any true causal claim relating c and e, the actual world events c and e must be causally related under the natural, or canonical, counterpart relation.

5.2.1 The Spectra of Counterpart Relations

For any given event c there exists a spectrum of potential counterpart relations. At either end of the spectrum are the extremal counterpart relations. At one end of this spectrum the event has no essence and so anything goes—the counterpart relation is so permissive as to allow any region of any other world to count as a counterpart of c. Call this *MAX* and any event c under that counterpart relation can be represented by c_{MAX}. At the other end of the spectrum every feature is essential and only a perfect intrinsic *and* extrinsic duplicate will qualify as a counterpart. A perfect intrinsic duplicate of c can occur in a different world, but a perfectly intrinsic and extrinsic duplicate can only occur in a world which is a perfect duplicate of the actual world in every way—that is not another possible world, that just *is* the actual world.[2] So under this maximally strict counterpart relation all claims about c in non-actual situations are false as c has no counterparts in non-actual worlds. I will call this counterpart relation *MIN* and represent event c under *MIN* as follows: c_{MIN}.

Plugging actual world event c into Lewis's original causal conditional under either of these extreme counterpart relations does not yield substantially informative propositions:[3]

L1 $\neg Oc_{MAX} \mathrel{\Box\!\!\rightarrow} \neg Oe$ is always true for any e, since the antecedent is trivially false—there are no $\neg Oc_{MAX}$-worlds, worlds where c_{MAX} does not occur.

L2 $\neg Oc \mathrel{\Box\!\!\rightarrow} \neg Oe_{MAX}$ is always false, unless there are no $\neg Oc$-worlds (i.e unless c occurs in every world), since the consequent is trivially false—there are no $\neg Oe_{MAX}$-worlds, worlds where e_{MAX} does not occur.

L3 $\neg Oc \mathrel{\Box\!\!\rightarrow} \neg Oe_{MIN}$ is always true. Since e_{MIN} only occurs in the actual world, the consequent $\neg Oe_{MIN}$ is true in every non-actual world. Where the consequent is always true, the conditional is always true.

L4 $\neg Oc_{MIN} \mathrel{\Box\!\!\rightarrow} \neg Oe$ is a little more complicated. Since c_{MIN} only occurs in the actual world, the antecedent is true in every non-actual world. This means

[2]This assumes that no two distinct worlds are qualitatively identical. I think this is a reasonable and simplifying assumption but it is not strictly necessary. Even if there are distinct worlds which are qualitatively identical, and so there are worlds where c occurs under this minimal counterpart relation, conditionals which take c in the antecedent place, will never be substantially informative.

[3]For the purposes of illustration I revert temporarily to Lewis's formulation where the cause side and the effect side are subject to the same (somewhat ambiguous) standard of non-occurrence: $\neg Oc$ and $\neg Oe$.

that the conditional is false unless there is some $\neg Oe$-world which is closer than every Oe-world. It is hard to imagine a case where the counterpart relation is such as to render every $\neg Oe$-world closer than every Oe-world and where the L4 counterfactual is substantially informative. For example, if e was taken to fall under the *MIN* counterpart relation too, then every $\neg Oe$-world would be closer than any Oe-worlds (since there are no Oe-worlds other than the actual). However, if both Oc_{MIN} and Oe_{MIN} occur only in the actual world, then every causal counterfactual that takes their negation in the antecedent and consequent place will be trivially true.

If events were subject to such extremal counterpart relations, and if a Lewisian counterfactual analysis of causation is remotely plausible, then they would be caused by everything (or nothing), or would cause everything (or nothing) irrespective of the physics of the world that c and e occupied. This cannot be right. Given that Lewis's counterfactual analysis of causation is antecedently plausible, I conclude that the problem lies with the extremal counterpart relations. These cannot feature in a meaningful and substantial causal test.

Interestingly, the result is the same (albeit for slightly different reasons than in L4) when we introduce the asymmetric standard of non-occurrence that I introduced in the last chapter (the ACCT test for causal dependence):

ACCT 1 $\neg Pc_{MAX} \boxdot\!\!\rightarrow \neg Oe$ is always true for any e, since the antecedent is trivially false—there are no $\neg Pc_{MAX}$-worlds.

ACCT 2 $\neg Pc \boxdot\!\!\rightarrow \neg Oe_{MAX}$ is always false, unless there are no $\neg Pc$-worlds (i.e unless c partially occurs in every world), since the consequent is trivially false—there are no $\neg Oe_{MAX}$-worlds.

ACCT 3 $\neg Pc \boxdot\!\!\rightarrow \neg Oe_{MIN}$ is always true: since e_{MIN} only fully occurs in the actual world, the consequent $\neg Oe_{MIN}$ is true in every non-actual world. Where the consequent is always true, the conditional is always true.

ACCT 4 $\neg Pc_{MIN} \boxdot\!\!\rightarrow \neg Oe$ is always true. Recall that c_{MIN} is c under a counterpart relation in which every intrinsic feature of c and every extrinsic feature of c's world (the actual world) are essential. For c to partially occur under that standard of counterpart relation merely requires that a single essential feature remains, so for c not to even partially occur ($\neg Pc_{MIN}$) would require that there be an alteration of c in a world where no feature of the actual world remained. Since e occurs in the actual world, every $\neg Pc_{MIN}$-world must be a world where no feature of e remains and so $\neg Oe$ must be true. So, in all the closest $\neg Pc_{MIN}$-worlds, $\neg Oe$ must be true rendering the conditional true regardless of what c and e pick out.

Again, the extremal counterpart relations make a nonsense of an other-wise plausible theory. These counterpart relations cannot apply in our counterfactual causal reasoning.

Notice, though, that the spectrum of counterpart relations splits into two sub-spectra: there are those counterpart relations whose members stand in that relation in virtue of some intrinsic property or combination of intrinsic properties (e.g. the counterpart events are those that contain an object which is round and metal), and those counterpart relations whose members stand in a counterpart relation at least partly in virtue of extrinsic properties (e.g. where all counterparts take place north of the equator) or who stand in that relation regardless of their properties (e.g. where the only counterparts are the laying of an egg, a performance of Macbeth and the spinning of a single electron). I will refer to the first sort of relation as 'natural' and the second as 'arbitrary'. The sub-spectrum made up of only the arbitrary counterpart relations has the same two extremal relations as the original spectrum of all counterpart relations, *MAX* and *MIN*. However, because *MIN* relates only those intrinsic and *extrinsic* duplicates, it does not feature on the sub-spectrum made up of only the natural counterpart relations, since the natural relations exclusively concern the intrinsic features. That means that the strictest natural counterpart relation there can be is one which relates the intrinsic duplicates regardless of their extrinsic similarities or dissimilarities. I represent this counterpart relation by adding subscript i, so c under counterpart relation i is written as c_i.

This counterpart relation does not trivialise our causal test since it is a substantial question whether the following conditionals are true or false:

ACCT 5 $\neg Pc_i \mathrel{\Box\!\!\rightarrow} \neg Oe$. This conditional is true iff all of the closest worlds where the best candidate counterpart of c does not share a single intrinsic feature of c are worlds where the best candidate counterpart for e does not share all of e's essential features.

ACCT 6 $\neg Pc \mathrel{\Box\!\!\rightarrow} \neg Oe_i$. This conditional is true iff all of the closest worlds where the best candidate counterpart of c does not share a single essential feature with c are worlds where the best candidate counterpart for e does not share all of e's intrinsic features.

These conditionals are substantial because we must know more about the nature of c and e to know if they are true or not. This was not the case with the extremal counterpart relations. What is more, this counterpart relation seems to have a certain logical priority as it is the strictest counterpart relation that can be formed on the basis of intrinsic properties.

5.2.2 The Canonical Counterpart Relation

I think that there is an independent reason to embrace one particular non-extremal counterpart relation as canonical in our ontology. This conviction is based in part on the Humean assumption that there are no necessary connections between distinct entities. When giving our total description of an object, x, we will include intrinsic features, such as *is round*, and extrinsic features, such as *is on earth*. This total description picks out x but does so in reference to distinct entities such as earth (or any other object mentioned in its extrinsic feature set). If this total description entered our ontological canon, x

would be defined, at least in part, in reference to these distinct entities—it could no longer be x without them. This would establish a necessary connection between x and the distinct entities in x's world and so would violate the Humean assumption. The nature of an object would no longer just be a 'local matter of particular fact' as the Humean picture requires, according to Lewis [1986d, ix], but a matter of the fact concerning local and non-local parts of a system.

Respecting the Humean assumption requires that the canonical description of the object is that description which details only its intrinsic features. Reference to extrinsic features is useful and entirely unproblematic for our ordinary discourse, but when it comes to detailing our fundamental ontology and giving the most basic and fundamental (i.e. canonical) description of the world, then such descriptions should be eschewed.

Objects and events are analogous in this respect[4]—the canonical description of an event should detail only its intrinsic features lest it violate the Humean assumption. Having eschewed the extrinsic features of the event, the maximally strict counterpart relation that it can now fall under would require *all* of its intrinsic duplicates to be present in *every* counterpart.

Furthermore, the canonical description is neutral in respect of background grouping and so none of the intrinsic features are more or less important than any other. In the context of a canonical description, then, either all of the features are essential or none are. If none of the features were essential, then the counterpart relation in the canonical context would be equivalent to *MAX*, which, as we saw above (5.2.1 and 5.2.1), led to triviality in causal claims in particular, but also makes a nonsense of our modal attributions in general since every event would have a counterpart in every world rendering every event modally necessary! That cannot be. However, if all of the natural intrinsic features of the event are essential, then our causal and modal attributions make sense, and the Humean assumption is not violated. So, the counterpart relation invoked by the canonical description of an event should treat all, rather than none, of the intrinsic properties as essential. This is exactly what the i counterpart relation introduced above does. I take this to motivate thinking that the i counterpart relation is the *canonical* counterpart relation for events.

Some will object that without context there can be no sensible discussion of which counterpart relation applies to an object or event. I am inclined to think otherwise but I need not argue for this position here. It will suffice for this purpose that in the context of a canonical description, there is a canonical counterpart relation that applies. The context in which the canonical description applies is an idealised context and resides firmly outwith our practical reach. Nevertheless it is the context in which the object or event get its definitive place in our inventory of the world—it is the *privileged context*.

All that this demonstrates is that there is some some stand-out counterpart relation (the strictest) on the spectrum which is non-arbitrary and which does not trivialise counterfactual conditionals concerning that event. What remains to be shown is that this counterpart relation has some important relation to our causal talk and that it is

[4]This is a consequence of my ontology of events, where events are taken to be space-time regions of a world—see Chapter 2.

useful in resolving the contextualist-naturalist tension.

5.2.3 Canonical Implications

Before going on to argue that the canonical counterpart relation does indeed help resolve the contextualist-naturalist tension within causation, I will first consider some of the implications that follow from adopting this counterpart relation for events within a counterfactual theory of causation.

Each version of counterfactual analysis I have considered so far has agreed on the following: c is a cause of e if and only if there exists a chain of causal dependence between c and e. I have argued for certain refinements of what determines causal dependence, i.e. which counterfactual conditional determines causal dependence. At the end of Chapter 4 I settled on an asymmetric, counterpart-theoretic, counterfactual theory of that dependence (ACCT) which takes e to causally depend on c relative to a counterpart relation x iff $\neg Pc_x \boxdot\!\!\rightarrow \neg Oe_x$. In the privileged context, the counterpart relation in question is the canonical counterpart relation (i.e. the variable x is set to i). So, in the privileged context, e causally depends on c iff: $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$. Call this test for causal dependence in the privileged context **ACCT Canonical**, or Canonical for short.

Now, let us suppose for the sake of discussion that the test for causal dependence in the privileged context (Canonical) is the definitive test for causal dependence between distinct events in our theory of causation. According to our present physics, this would provide an extremely permissive test for causation indeed. Within relativistic physics, every event in the backwards light cone of an effect will have had at the very least a gravitational influence on that effect. Cleanly excise any past event relative to the effect and the effect will be altered in some way, however minute. Therefore, by adopting Canonical as part of our definitive causal test, every event in the backwards light cone of an effect is a cause of that effect.[5]

Of course, had our physics been different, this may not have been the case. Endorsing Canonical as part of our definitive causal test does not directly result in the claim that every event in the past of some effect is a cause of that effect, but it does give that result when the physics tells us that this is how events influence one another. This is a positive result for three reasons: First, Canonical is sensitive to what the physics of the world is, surely this is a pre-requisite for a viable theory of causation. Second, assume it were otherwise and our causal theory was not sufficiently sensitive so as to attribute causation where our physics can detect an influence. A pedantic physicist could claim to have discovered and been able to manipulate causal connections that our causal theory could not endorse—a disaster for that causal theory. Third, and most importantly, we know from the chaotic nature of our world that minute differences in initial conditions can yield enormous differences elsewhere in the system. If we were to ignore some variables at the initial conditions, we could never tell the complete causal story of some later macro-level event. So, our causal theory needs to be able to accommodate *all* of the features of the initial conditions. Adopting Canonical as our definitive test

[5]This point is familiar from Latham [1987], Field [2003] and Schaffer [2005].

for causal dependence avoids these issues by endorsing every claim of influence that our physical theory can support as causal. Of course, that makes for a very great deal of causation in our world. It is simply too much for some.[6]

I take it this is what Lewis had in mind when he rejected a 'uniformly stringent' standard of event fragility [1986d, p.199]. Lewis was concerned that a test such as Canonical let in too much: it makes the bystander, the pre-empted back-up (Billy), the dog barking several streets away and the gravitational influence of Jupiter all count as causes of the window breaking event. By adopting the Canonical test within our causal theory, however, they *are* all considered causes, as is the firing of the gun last year, the Great Fire of London or the movement of the remotest dust mote on Mars. Absolutely everything in the backwards light cone is a cause of the death. And yet we discriminate between the guilty and the innocent, the vandal and the bystander, the person who fired the gun and the person who tried to stop them, the relevant factors from the background conditions and so on. We discriminate as a matter of course but a uniformly strict standard of fragility, i.e. the standard applied in Canonical, cannot. Lewis thought this a *reductio* of the fragile view of events (and, we may presume, the Canonical test by extension). I do not.

We do discriminate between causes and non-causes and between background and foreground conditions for an effect, this much is obvious from our causal attributions, but why should that make us think that there is such a discrimination at the level of metaphysics? Recall Putnam's Venusians [1982] arriving at earth and noting all of the combustible oxygen lying around. They may well think that it was the oxygen that caused the fire, not the spark that we humans typically blame. Or take the normally oxygen-free lab where sparks are common. The oxygen we typically consider a background condition for ordinary fires is suddenly taken to be the cause of the fire in the lab. The selections we impose when making causal attributions are influenced by a wide range of contextual factors, and these are surely relevant in analysing our causal talk and our causal assertions, but if there is a mind-independent, objective matter of fact that we track with those assertions, we had better look for a fact which is neutral with respect to such selections. We need a pre-selective account of the metaphysical relation of causation quite irrespective of the obviously selective nature of our causal talk.

Here we reach a methodological fork in the road. Either we treat our causal assertions as transparently manifesting a context sensitive, selective, notion to be analysed and work on that or, instead, we take our causal assertions to imperfectly track some objective feature of the world and seek an understanding of that feature. The former path is well trodden: Hitchcock & Knobe [2009], McGrath [2005], Menzies [2009] and Schaffer [2000b] can all be seen to be following this methodological road. The latter path is something of a tightrope as it requires using causal assertions and intuitions at once as evidence and at another time as data to be explained away. But which are the good examples of causal talk and which are the bad? Which are good evidence and which are to be explained away? I believe it is worries of this sort that make the first

[6]Objectors include Paul [2000, p.236], Lewis [2004a, p.88-90], Schaffer [2005, p.334] and Menzies [2009, p.349-351].

path so popular by comparison and the second so lonely.[7] Nevertheless, I think the latter path remains worth exploring. If there is a mind-independent, objective feature of the world that our causal assertions track, the first path will never lead us to it. It is antecedently plausible that such a feature of our world exists so it is reasonable to start with the hypothesis that such a feature exists and see if the data can be understood in light of that starting assumption. If it cannot, or if the concessions required to make the data fit are just too much then we will have made progress, albeit of a negative sort. If it can, and if the explanations for the deviant data are plausible then that leaves our central causal platitude intact. I think that that is a balancing act worth attempting.

So, revisiting Lewis's objection, we see the complaint is that the fragilicist's metaphysics does not map *directly* onto our causal talk. I would prefer to put it around the other way: our causal talk does not map directly onto the metaphysics being proposed. Put that way it is no surprise. Our assertions about objects are loose and vague in a way that our physics most certainly is not. We talk of windows breaking and fires starting and these macro-level assertions do not precisely map on to the micro-level facts about atoms and molecules, let alone electrons, fields and wave functions. We can reasonably (though not uncontroversially—see Anderson [1972] and Batterman [2013]) suppose that with a dose of charitable interpretation our assertions about the macro level will meaningfully relate to the micro-level facts and it is an interesting project to investigate what that relation might be. However, along the way we are bound to come across acceptable, yet strictly false, claims about matter being solid (all matter is very nearly entirely made up of empty space) or wood not conducting electricity (it does, just very little) that can be charitably read as being true relevant to context-given values for solidity or conductivity. It is no threat to our scientific theory that such divergence from common talk exists but it does highlight a double standard: a canonical fact of the matter about a value, and a contextually variant fact about the relevance of that value. What is interesting to note is that for the canonical fact of the matter to be truly canonical, it must be consistent with the widest range of contexts in which it is invoked, but for any non-canonical claim made concerning that fact there can be no certainty that the claim will hold in another context. Take the conductivity of wood, for example. The canonical fact (as I am calling it) is that wood does conduct electricity. The person that says it does not may be saying something useful or interesting in their context but there still remains another context—the physics lab perhaps—where the conductivity of wood is known and discussed. The canonical fact that wood conducts at least some electricity must remain stable across these contexts whereas the contextually embedded claims need not. This is an important asymmetry.

Of course, it would be a grievous error to let the tail wag the dog here. Our common sense assertions are held to account if they deviate from scientific findings, not the other way around. Philosophical theorising cannot hope to take such epistemic precedence over common sense, but if there is a mind-independent, objective truth of the matter

[7] To my knowledge Hall and Paul [2013] are the only philosophers in recent years to have take seriously the notion of a reductive account of causation, and even they are not persuaded that it is viable.

about what causes what then the lesson from scientific progress strongly suggests that we will only loosely approximate the truth in our ordinary talk. Taking such talk too seriously begs the question against the naturalist. However, if the naturalist can show that our causal talk tracks some mind-independent feature of the world, even loosely, then there remains hope for their position.

I will offer an analysis of causation which would make causation a natural relation between distinct events (naturalism). I will show that our causal talk tracks this natural relation between events despite varying in truth across contexts (contextualism). The resultant view will give a two-tiered test for causation. The first test will determine the natural fact concerning the causal relation of c and e, the second will determine the context-sensitive truth of a causal assertion concerning c and e.

I am now in a position to outline my central thesis:

> **Thesis 1**: There is a mind-independent fact of the matter about whether one event c caused another e simpliciter: e causally depends on c relative to counterpart relation x iff $\neg Pc_x \boxdot\!\!\rightarrow \neg Oe_x$; and c is a cause of e *simpliciter* iff there is a chain of counterfactual dependence between c and e under the canonical counterpart relation i (i.e. between c_i and e_i). This is my analysis of the causal relation.
>
> **Thesis 2**: Our causal talk must at the very least track this fact of the matter about the causal relation if it is to express literally true causal claims. However, that does not mean that the only true causal claims are those which directly express such facts.

Thesis 1 is simply Lewis's original counterfactual analysis, applied in the canonical context and with a disambiguated asymmetrical standard of non-occurrence applied to c and e. I have argued for the stand-out status of the canonical context in this chapter and for the double-standard of non-occurrence in Chapter 4 and, by plugging these refinements into Lewis's counterfactual analysis, I have proposed an extremely permissive, mind- and interest-independent, analysis of the causal relation. At the heart of this analysis of the causal relation is the ACCT Canonical test for causal dependence: e causally depends on c relative to the canonical counterpart relation iff $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$ is true.

As I have already noted, however, our causal talk is not so permissive and does not take place in the privileged context. Our causal talk operates in contexts where the i counterpart relation does not apply and so we should not expect our causal talk to be subject to the analysis offered for the causal relation itself. In Chapters 2, 3 and 4 I argued for a contextualist understanding of our causal talk that was subject to variations in the counterpart relations of the events in question. That contextualist argument held that the claim 'c is a cause of e' is true iff c and e in that context (i.e. c_n and e_n) were linked by a chain of causal dependence, where causal dependence was determined by the truth of a context-sensitive counterfactual: $\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n$. At the heart of this analysis of our ordinary causal talk is the ACCT Contextual test for causal dependence: e causally depends on c relative to the contextually determined counterpart relation n iff $\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n$ is true.

So, I have two analyses on offer: one for our ordinary causal talk (featuring ACCT Contextual) and the other for the causal relation itself (featuring ACCT Canonical). They share a common core in that each requires that there be chains of causal dependence between c and e events, and they even agree on the form of the counterfactual dependence relation that must hold if there is to be causal dependence: $\neg Pc_x \; \Box\!\!\rightarrow \neg Oe_x$. Applied to our causal talk the counterpart place (x) takes the value n as given by the ordinary context. In the context of questions about the causal relation in our world, the counterpart place takes the value i, as is mandated by the privileged context. The first fits the contextualist data discussed in Chapter 2 and the second fits the 'central causal platitude' that whether c and e are causally related is a mind-independent, objective, matter of fact. I will refer to this two-tier application of the asymmetrical counterpart-theoretic counterfactual theory of causal dependence as Double-ACCT.

The purpose of the remainder of this chapter is to show that these two analyses are compatible and how they relate. Each analysis differs only on standards of causal dependence on offer so I will focus on how those standards—ACCT Canonical and ACCT Contextual—relate. I think it is obvious that the two analyses are aiming at different goals—one aims to analyse the causal relation, the other our causal talk—and so there is no deep conflict here, but it is important to understand how these two notions relate. If there is no relation between the two then there are two topics being conflated in my discussion—it would amount to a verbal dispute about what the topic of concern was: the causal relation or causal talk. This would be particularly problematic for my view as I am using some of the data from our causal talk to inform my account of the causal relation. They had better not be different topics!

On the other hand, it is antecedently plausible that there be a relation between the truth of our causal assertions and the causal truths of the world (assuming there are any, which I do), so it would be good for my theory if it explained and articulated that relation. For example if it were the case that $\neg Pc_n \; \Box\!\!\rightarrow \neg Oe_n$ were true only if $\neg Pc_i \; \Box\!\!\rightarrow \neg Oe_i$ were true, then that would show that, within Double-ACCT, for any given causal assertion to be true there must be a corresponding causal relation in the world that made it true. That would be a satisfying result, indicating the fundamentality of the causal relation and our success in tracking that relation in our ordinary causal talk. Things are not quite so simple, however, as I aim to explain in the next section.

5.3 Causal Talk

In this section I will show that, with an important exception, if a causal assertion is true under in some ordinary context, i.e. if $\neg Pc_n \; \Box\!\!\rightarrow \neg Oe_n$ is true, then it is true in the privileged context, i.e. $\neg Pc_i \; \Box\!\!\rightarrow \neg Oe_i$ will be true. The exception in question will ultimately motivate an addendum to my analysis of causal talk. First I will introduce the terms and notation I will be using, then I will argue that $\neg Pc_n \; \Box\!\!\rightarrow \neg Oe_n$ is true only if $\neg Pc_i \; \Box\!\!\rightarrow \neg Oe_i$ is true. I will then introduce a counter-example to that argument and derive a crucial lesson from it.

5.3.1 Terminology

I will first introduce some terminology and notation that I will be working with. Some of it will be familiar already but here I state the terms for clarity.

- $x \mathrel{\Box\!\!\rightarrow} y$ is true iff in all of the closest possible worlds where the proposition x is true, the proposition y is true.[8]

- The variables c and e range over events. I take events to be regions of spacetime of a world and I will use c and e respectively to denote the (putative) *cause* and *effect*.

- I will take the *counterparts* of c and e to be regions of worlds. The regions occupied by c and e in the actual world are automatically counterparts of c and e respectively and no other region of the actual world is. Which regions of non-actual worlds qualify as *counterparts* of c and e is determined by a similarity-metric. This similarity metric is determined, at least in part, by context and by how c and e are represented. This makes counterparthood a mind-dependent matter. Reference to *candidate counterparts* will indicate that we are to remain neutral on which standard of similarity applies in the context in question. This will be useful when comparing a range of prospective similarity metrics within a context.

- The *counterpart relation* is that relation which holds between individuals and their counterparts. Given the context-sensitivity of whether a given individual is a counterpart of another or not, the counterpart relation varies with context. I have argued in this chapter that in the privileged context, an individual c shares all and only its intrinsic features with every one of its counterparts. I have referred to this as the *canonical counterpart relation* for c. There is a weaker similarity metric, where c shares at least *some* particular intrinsic feature or features with all of its counterparts. When that is the case, the counterpart relation that applies in that context can be said to be *natural* with respect to c.[9] Notice that the canonical counterpart relation is a special case of a natural counterpart relation and that those features which are essential under a natural counterpart relation—I'll call them n-features of c—are a proper subset of the features which are essential under the canonical counterpart relation—I'll call them i-features of c.

- I will use c_i to refer to c under the *canonical* counterpart relation. This counterpart relation is determined by the i-features (complete set of intrinsic features

[8] As previously, I am simplifying the truth conditions here is a way that implies the Limit Assumption. I believe this is harmless in this context.

[9] This is to be contrasted with unnatural or arbitrary standards of similarity between c and its counterparts. Such standards of similarity take non-actual individuals to be counterparts of an actual individual on the basis of some non-intrinsic features of c, or which base the similarity on no common features of the set of counterparts at all.

plus spatiotemporal location) of c. I will use c_n to refer to c under a *natural* counterpart relation. This counterpart relation is fixed by which i-features (intrinsic features plus spatiotemporal location) of c are taken to be essential in a given context.

- I argued in Chapter 4 that there are two distinct standards which we might invoke when we consider the negation of 'c occurred'. The first standard of non-occurrence I refer to as *clean excision* and this is a particularly strong standard of non-occurrence of an event. On this standard, to say that c did not occur in a world means that by the standards of similarity that apply in that context, there is no candidate counterpart of c in that world which shares even a single essential feature of c. In other words c does not even partly occur in that world relative to a given counterpart relation. The set of worlds which meet clean excision standard of non-occurrence for c will be represented as $\neg Pc_x$ where x indicates the counterpart relation. The second standard of non-occurrence I refer to as closest-alterative. On this standard, to say that e did not occur in a world means that there is no candidate counterpart of e in that world which shares *all* of the essential features of e. In other words e does not fully occur in that world. The set of worlds which meet closest alternative standard of non-occurrence for e will be represented as $\neg Oe_x$ where x again indicates the counterpart relation.

I have introduced a two-tier thesis: one test for our causal talk and another for the relation of causation itself. Both share the following ACCT element: e causally depends on c relative to counterpart relation x, iff $\neg Pc_x \mathrel{\Box\!\!\rightarrow} \neg Oe_x$ is true. My account of our causal talk sets x to n, where n is a function of context and representation. My account of the causal relation sets x to i, where i is the canonical counterpart relation as mandated by the privileged context. Now I can express the truth conditions for the two tests for causal dependence, which sit at the heart of the analyses of causal talk and the causal relation respectively, using the terminology and notation just introduced:

- ACCT Contextual: $\neg Pc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_n$ is true iff in all of the closest possible worlds in which no counterpart of c with even a single n-feature occurs of c, no counterpart of e with *every* n-feature of e occurs.
- ACCT Canonical: $\neg Pc_i \mathrel{\Box\!\!\rightarrow} \neg Oe_i$ is true iff in all of the closest possible worlds in which no counterpart of c with even a single i-feature of c occurs, no counterpart of e with *every* i-feature of e occurs.

5.3.2 The Reduction Argument

Having established the terminology and notation in the previous section, and having re-introduced the ACCT Contextual and ACCT Canonical tests for causal dependence, I am now in a position to argue for the target proposition: if e causally depends on c under any ordinary counterpart relation then e causally depends on c under the canonical counterpart relation too.

First, notice that if e does not fully occur in some world by the standards ordinary context, it does not fully occur in that world by the standards of the privileged context.

1. Suppose a natural counterpart relation n is given. If a candidate counterpart of e lacks even one of the features that are essential given that relation—if $\neg Oe_n$ is true—then that candidate counterpart lacks one of the i-features. Therefore, any world in which $\neg Oe_n$ is true, $\neg Oe_i$ is true. $\neg Oe_n$ entails $\neg Oe_i$.

   Example If throwing the heavy red ball did not wholly occur, then nor did the throwing of the heavy, red, shiny, leather...etc. ball.[10]

Second, notice that worlds in which c has been cleanly excised relative to to an ordinary context are at least as close to the actual world as worlds in which c has been cleanly excised relative to the privileged context:

2. Those possible worlds in which c_n is cleanly excised lack some of the i-features of c. Those possible worlds in which c_i is cleanly excised lack all of the i-features of c. Therefore, all else being equal, the closest $\neg Pc_n$-worlds will be closer than the closest $\neg Pc_i$-worlds, except where n is fixed by all of the i-features in which case they are equally close. So the closest $\neg Pc_n$-worlds are at least as close as the closest $\neg Pc_i$-worlds.

   Example All else being equal, the worlds where you throw nothing (no heavy thing, no red thing, no ball) are further away than the worlds where you throw something (a heavy rock) similar to that which you threw in the actual world (a heavy red ball).

Third, suppose that $\neg Pc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_n$ is true. By 1, this entails that $\neg Pc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_i$. For clarity:

3. $\neg Pc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_n$ entails $\neg Pc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_i$

   Example If you hadn't thrown the rock then the window would not have broken therefore if you hadn't thrown the rock the window would not have broken exactly as it did (into 357 pieces, at such-and-such a time... etc.)

Fourth, we are supposing that $\neg Pc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_n$ is true and so by 3, we can infer that $\neg Pc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_i$ is true too. For $\neg Oe_i$ to be true in a world requires that the candidate counterpart for e in that world be unlike the actual e in at least one i-feature. If the closest of the $\neg Pc_n$-worlds are sufficiently unlike the actual world to make $\neg Oe_i$ true, then (from 2) the no-closer $\neg Pc_i$-worlds will be sufficiently unlike the actual world to make $\neg Oe_i$ true there as well. Thus in all of the closest of the $\neg Pc_i$-worlds, $\neg Oe_i$ is true. So:

4. $\neg Pc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_i$ entails $\neg Pc_i \mathrel{\Box\!\!\rightarrow} \neg Oe_i$.

[10] It will help to indicate a list of all of the intrinsic features of an event with an ellipsis and an 'etc.'.

> **Example** If you hadn't thrown the rock, the window would not have broken into 357 pieces at such-and-such a time... etc., therefore if you hadn't thrown the rock in just that way at just that time... etc., then the window would not have broken into 357 pieces at such-and-such a time... etc.

So, on the assumption that $\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n$ is true, we can derive that $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$ is true. If correct, this would mean that any causal claim that passes the ACCT test for causal dependence in an ordinary context (ACCT Contextual), will also pass the same ACCT test in the canonical context (ACCT Canonical). Relating this back to the analyses offered of our causal talk and of the causal relation in the Double-ACCT theory, this result suggests that if it is true to *say* that c is a cause of e, then it must be true that c and e are causally related. This indicates that our causal talk asymmetrically depends on the way the world is. A neat and plausible result.

5.3.3 A Problem for Reduction

There is a problem in the foregoing argument. The argument for 4 states:

> If the closest of the $\neg Pc_n$-worlds are sufficiently unlike the actual world to make $\neg Oe_i$ true, then (from 2) the no-closer $\neg Pc_i$-worlds will be sufficiently unlike the actual world to make $\neg Oe_i$ true there as well.

However, this inference is fallacious. It is an instance of strengthening the antecedent which Lewis discusses in *Counterfactuals* [2001, p.32]. When we adopt a new, stronger, antecedent we alter where the closest antecedent-satisfying worlds are: they are now further away since a greater alteration has had to take place to satisfy the more-demanding proposition that $\neg Pc_i$ occurs. The inference would be valid if we could guarantee that a more distant world ensured the consequent remained true ($\neg Oe_i$), but we simply cannot guarantee that: it may be the case that as the closest antecedent-satisfying worlds become more dissimilar overall, they at some point become more similar in respect of the region associated with the consequent (the region corresponding to e in the actual world). Whilst it is not guaranteed, the cases where it fails will be quite specific: for $\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_i$ to be true and $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$ to be false, requires that the shift from $\neg Pc_n$ to $\neg Pc_i$ is attended by a shift from $\neg Oe_i$ to Oe_i. What would such a case look like?

> **Gun-Balloon**: Taking aim at an escaped balloon, a child squeezes the trigger on a toy gun. The safety catch is on and the bearing does not fire. The balloon floats away on the wind.

Let us signify the event where the child squeezes the trigger as c and the balloon floating off as e. Does c cause e? Adopting the ACCT Canonical test for causal dependence (and ignoring their gravitational influence on one another just for the sake of illustration), we start by assessing the truth of the following counterfactual: $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$. Since absent the c_i event (i.e. in the closest worlds where $\neg Pc_i$ is true) the balloon still floats off (Oe_i is true), e does not causally depend on e in the privileged

context and so c is not a cause of e simpliciter. My contention in this discussion has been that for every true claim of the form 'c caused e' (where the context determines the counterpart relation n), it would be the case that c was a cause of e simpliciter, i.e. in the privileged context. So, if the latter is false (as it seems to be in the gun/balloon example above), there had better be no claim of the form 'c caused e' pertaining to those same events which is true.

In the example given, however, such a claim seems to be true by the Double-ACCT theory: where we take the safety catch being on as the only essential feature of c then c does appear to be a cause of e since all of the closest worlds where the safety is off ($\neg Pc_n$ is true) the trigger is still squeezed, the bearing fires, and the balloon is popped ($\neg Oe_n$ is true). So, for that standard of n, $\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n$ is true but $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$ is false, refuting my original claim that the truth of the first entailed the truth of the second. Again, assuming my two-part causal analysis, we cannot simply move from the truth of a causal claim in one context, analysed in terms of the $\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n$ condition, to their being a causal connection between the two events involved in the canonical context ($\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$). My neat and plausible result was wrong.

5.4 A New Causal Analysis

5.4.1 A New Causal Test

This problem example is interesting however because it is a case of prevention, a particular example of the broader category of *absence* causal claims. I will discuss absence causal claims in much more detail in the next chapter but for now it will suffice to say that I am not alone in being sceptical that we should endorse absence claims as literally true causal claims, despite their widespread use in our causal talk.[11] When we say that the rain prevented the fire, I think we give a causal explanation without the rain and fire ever having any *actual* connection. They cannot have an actual connection, of course, since the fire did not actualise!

Will all exceptions to my neat result have this absence structure? Not quite, but I think they will all demonstrate something very like it. The problem cases arise out of there being no actual connection between the cause event and the effect in the canonical context—that is to say that the clean excision of the cause will have no impact on the effect—but where a specified alteration to the cause will bring about some alteration in the effect. The specified alteration in the cause concerns a non-actual configuration of the features of the cause event and so it is a supposition about how the cause might have been different. The clean excision of *that* version of the cause would make a difference to the effect and so that would-be event should be considered a cause of that would-be effect. All of this is at a remove from the actual occurrences of our world, as it must be to create the problem in the first place. Put that way, it should be clear that would-be causation is not a case of genuine causation, as genuine causation relates actual-world events and would-be causation relates possible-world events. I will argue in the next chapter that all cases of absence causation are cases of would-be causation

[11] See Beebee [2004], Dowe [2004b] and Schaffer [2004a] for example.

and should not be considered examples of genuine or actual causation. That argument remains to be made, however, so for the time being I will consider the problem cases to split between would-be causal claims (which are manifestly not genuine causal claims about our world) and absence causal claims about which I appeal to existing scepticism in the literature.

Taking this scepticism about absence causal claims at face value for the moment, it suggests that ACCT Canonical test ($\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$) gave us the right answer about the causal connectedness of the two events. That is in line with my analysis of the causal relation as given in §5.2.3.

> **Thesis 1**: There is a mind-independent fact of the matter about whether one event c caused another e simpliciter: e causally depends on c relative to counterpart relation x iff $\neg Pc_x \boxdot\!\!\rightarrow \neg Oe_x$; and c is a cause of e *simpliciter* iff there is a chain of counterfactual dependence between c and e under the canonical counterpart relation i (i.e. between c_i and e_i). This is my analysis of the causal relation.

The second part of the analysis outline then concerned the connection between this natural fact and the assertions of our causal talk:

> **Thesis 2**: Our causal talk must at the very least track this fact of the matter about the causal relation if it is to express literally true causal claims. However, that does not mean that the only true causal claims are those which directly express such facts.

In section 5.3.2 I attempted to argue that if $\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n$ were true then $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$ would be true too. The idea was to show that if a causal claim is true in one context (i.e. with counterpart relation n) then that same claim would be true in the canonical context (i.e. with the same events under counterpart relation i). Were this correct it would have established a connection between the contextual and canonical causal claims being made just as Thesis 2 requires. However, it was not true as the problem case of prevention demonstrated. If we assume, as I do, that absence causal claims are literally false then the fact that $\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n$ was true in the prevention case demonstrates that this counterfactual test for our causal *assertions* (incorporating the ACCT Contextual test for causal dependence) has failed—it gave a true result for a false causal claim. So, we need a new test for the truth of our causal assertions.

Notice that the false result in the prevention case does not establish that the ACCT Contextual test ($\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n$) is irrelevant, just that it is not sufficient to determine the truth of the causal assertion. In order to rule out the false positives a further constraint is required, one which tests not only whether a certain modification of the cause would alter the effect, but if the cause as it *actually* was contributed to the effect. For this, I propose adding the following condition which I call **ACCT Actual**: e actually depends on c iff $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_n$. To meet this ACCT Actual condition it must be the case that in all of the closest possible worlds where the cause event is cleanly excised (including all of its intrinsic features, not just those taken to be essential-in-context) the candidate counterpart for the event lacks at least one context-sensitive

essential feature. This means that if the cause simply had not occurred, and if the effect would have remained (essentially) unchanged, the conditional would be false. Unlike the fully contextualised ACCT Contextual condition: $\neg Pc_n \mathrel{\Box\!\!\rightarrow} \neg Oe_n$, this new Actual condition will not give false positives in cases where *if* a specific feature of c had changed, then e would have been different. ACCT Contextual cannot tell actual from possible causal connections, but the new condition, ACCT Actual, can. That is why the new condition is required.

Can the new condition replace the old? No. As with the Canonical test, it simply lets in too much—anything in the backwards light cone that contributed to the effect having a certain essential feature will be a cause. Our causal talk is much more selective than this and it is our causal talk that this condition is supposed to help capture. Also, it will give false positives. Suppose a colour-blind bull is trained to charge at the waving of a certain embroidered rag but someone, not knowing this, said 'it was the red colour of the rag that caused the bull to charge'. On the ACCT Contextual test they would be taken to have spoken falsely as in all of the closest worlds where the red colour (essential feature) of the rag is absent, the bull still charges because the rag is still embroidered in the same way. On the ACCT Actual test, however, the clean excision of the rag-waving event, complete with all of its intrinsic features, means that there is no redness, no rag and no embroidery in the closest possible $\neg c$-worlds. Without the embroided rag the bull will not charge and so the new hybrid counterfactual test comes out true: $\neg Pc_i \mathrel{\Box\!\!\rightarrow} \neg Oe_n$. This erroneous result shows that the ACCT Actual test is not sufficient to establish the truth of a causal assertion. Both tests are required.

But what of the connection between this new two-part test for causal assertions—ACCT Contextual and ACCT Actual together— and that of the natural causal relation, ACCT Canonical? My outline analysis requires that the test for the causal talk track the test for the natural causal relation between the events. The new two-part test for causal assertions does track the natural relation in the following way: if some causal assertion concerning c and e is true on both ACCT Contextual and ACCT Actual tests, then that entails that c is a cause of e simpliciter. This follows simply from combining the new ACCT Actual test, which is satisfied when $\neg Pc_i \mathrel{\Box\!\!\rightarrow} \neg Oe_n$ is true, and the conclusion in §5.3.2, item 1, that if any world satisfies $\neg Oe_n$ then that world satisfies $\neg Oe_i$. So, if all of the closest $\neg Pc_i$-worlds are $\neg Oe_n$-worlds (as it must be for ACCT Actual to be satisfied), and all $\neg Oe_n$-worlds are $\neg Oe_i$-worlds, then all of the closest $\neg Pc_i$-worlds are $\neg Oe_i$-worlds (which satisfies ACCT Canonical). So, the truth of $\neg Pc_i \mathrel{\Box\!\!\rightarrow} \neg Oe_n$ (ACCT Actual) entails the truth of $\neg Pc_i \mathrel{\Box\!\!\rightarrow} \neg Oe_i$ (ACCT Canonical). Thus, any causal claim that is true in an ordinary context, is true in the privileged context. I take this to show that the truth of our causal assertions asymmetrically depends on the objective causal facts of our world.

5.4.2 The Final ACCT Analysis

I can now state my causal analysis. Given the common ACCT structure of the conditionals involved I will refer to this simply as the **ACCT Analysis**: actual event e causally depends on actual event c iff c and e are distinct events and the following conditional is true: $\neg Pc \mathrel{\Box\!\!\rightarrow} \neg Oe$; c is a cause of e *simpliciter* iff c_i and e_i are connected

by a chain of causal counterfactual dependence; a causal assertion of the form 'c caused e' is true in context C iff $e_n$is connected by a chain of counterfactual dependence to both c_n and c_i, where counterpart relation n is a function of the context C.

So, for any events c and e there are three tests to establish the causal connection between them:

1: ACCT Canonical Test $\neg Pc_i \ \Box\!\!\rightarrow \neg Oe_i$ (or chains thereof)

If this conditional (or chains thereof) is true for c and e, then c is a cause of e simpliciter.

2: ACCT Actual Test $\neg Pc_i \ \Box\!\!\rightarrow \neg Oe_n$ (or chains thereof)

If this conditional is true for c and e then c made some *specified* difference to e—i.e. the difference concerning the essential features in that context.

3: ACCT Contextual Test $\neg Pc_n \ \Box\!\!\rightarrow \neg Oe_n$ (or chains thereof)

This test establishes that a *specified* difference in c makes the specified difference to e—i.e. the differences concerning the essential features of c and e in that context.

Two events can only be causally connected if they pass the ACCT Canonical test or are linked by a chain of events that do. A causal assertion can only be true if the events involved pass both the ACCT Actual and ACCT Contextual tests (or are linked by a chain of events that do). One might wonder about performing a reduction here. Passing ACCT Actual entails passing ACCT Canonical (but not vice versa) so it is tempting to drop Canonical altogether. However, this would be to remove the important distinction between our causal talk and the causal facts of our world. What is more, it is possible to pass the Canonical test and fail the Actual test, as in the case of prevention. I intend to make use of this fact in my discussion in the remaining chapters.

5.5 Resolving the Tension

We are now in a position to see that the analysis that I am advocating is both contextualist and naturalist regarding causal claims. For any c and e there is a fact of the matter about whether c is an actual cause of e without reference to the context of utterance. That fact of the matter is determined by the counterfactual dependence of c upon e in the privileged context, under the canonical counterpart relation. Causal facts are context invariant, just as naturalism requires. In ordinary contexts, whether 'c is a cause of e' is true is determined by the counterpart relations that each is taken to have in the context of the causal claim being made. Thus, causal claims are context variant, just as contextualism requires.

It remains to be seen how the ACCT Analysis fares when applied to the problems of absence causation, transitivity and symmetrical redundant causation which beset extant counterfactual analyses. In the following chapters I will address each problem in turn and consider the prospects of my proposed counterpart-sensitive counterfactual analysis.

6

Absences, Prevention and Would-be Causation

Our causal talk is littered with examples of negative causes or omissions, as in when we say that the *absence* of oxygen caused the fire to go out, or when we say that the gardener's failure to water the plants caused them to die. It is incumbent upon a viable theory of causation to account for such talk of omissions. Whilst counterfactual theories of causation tend to be able to match intuition in the standard examples, three ontological worries emerge: First, where are these omissions supposed to *occur*? Second, if they occur at a distance from the effect doesn't this establish a highly controversial physical theory of action-at-a-distance too cheaply? Third: ordinary events such as my typing on this keyboard counterfactually depend on the omission of certain extraordinary events such as there being nerve gas or a velociraptor in the room with me. Are these ordinary omissions of extraordinary events causes of my typing? All of them?

The first of these problems I call *Location*, the second *Non-locality* and the third *Profligacy*. In §6.1 I will introduce and discuss each of these problems in turn before considering how certain existing treatments of absence causation handle them in §6.2. In §6.3 I introduce a novel approach and argue that it solves each of the problems neatly. I will then show that this approach coheres with my broader thesis.

6.1 Three Problems of Absence Causation

Why did the plants die whilst you were on holiday? It seems sensible to say they died because the gardener did not water them since, we suppose, if she had then they would not have died. Clearly "the gardener's omission caused the plants' death" is an acceptable assertion. An adequate account of causation should be able to accommodate

this datum. In this section I will discuss the problems of Location, Non-locality and Profligacy in turn. In the next section I offer a solution.

6.1.1 Location

At first glance a simple counterfactual analysis of causation appears to be well-equipped to accommodate absence examples. If there had been water, the plants would not have died, so the lack of water caused the plants to die. Absent the acid, the plants would have flowered, so the acid caused the plants not to flower, if there had been oxygen in the space, the fire would have continued to burn, and so on. The counterfactual apparatus seems to be able to give the right results by matching the truth of the counterfactuals to the causal intuitions in each case.

However, most counterfactual analyses are committed to an events ontology for the causal relata, where events correspond to regions of spacetime and so must have spatiotemporal location, but there appear to be at least two initial candidates for the location of *the gardener's not watering the plants*: it could be located at the plants where she *would have* watered them or it could be located where she was instead (on the couch having a nap). Without a definite location, an events-based analysis of causation cannot get started.

It gets worse, for there is quite a large window of time in which watering the plants would have averted their death and within that window any number of other ways in which the plant could have been watered and so specifying that the event took place in any given one of these precisely defined spatiotemporal locations would be arbitrary. Equally, however broad the window of time is within which the gardener could have successfully watered the plants then the period of her *failure* to do so is equally broad. No longer is the nap on the couch the obvious alternative to the location around the plants. Every location the gardener occupied during the many hours or days in which the plants could have been watered is now a candidate location for the omission of watering. Locating the omission is therefore a serious issue for a simple counterfactual analyses of causation involving events.

6.1.2 Non-Locality

Following Hall [2002], if we do suppose that the absence is located wherever the gardener is instead, then the gardener causes the plants to die at a distance—she need never disrupt the intervening space between herself and the plants to cause their death. Perhaps action at a distance is possible at our world, it is certainly conceivable, but physicists take this to be a substantial question about the physics of our world, not a trivial question answered by appeal to such everyday examples. Genuine action at a distance would demonstrate the *non-locality* of our physics and would contradict Special Relativity. Absences surely do not carry such weight for if they did, they would establish non-locality very cheaply indeed. We would have refuted relativity from the armchair.

Perhaps this is reason to locate the absence of water at the would-be location of the watering at the plants rather than at the gardener. However, this does not help in

prevention cases such as the following from Hall[1]:

Prevention Bomber is heading to Target but Fighter intercedes and shoots down Bomber. Target survives.

Here Fighter prevented the destruction of Target but did so at a spatiotemporal distance. We can suppose that Fighter has never been to Target, and has never disrupted the intervening space between Target and where Bomber was downed. If we take the failure to bomb to be located at Target where the bomb *would have* been dropped, then that establishes action at a distance on the cheap since the cause (Fighter's interception) and the effect (Target's survival) are spatiotemporally disconnected. On the other hand if you locate the failure to bomb at Bomber's wreckage then you avoid positing action at a distance in this case since Fighter and the wreckage are spatiotemporally proximal.

So, to avoid the problem of non-locality we need to locate the omission at the *actual* location of some positive event that takes place in some cases (e.g. the gardner example) and we need to locate the omission in the *would-be* location of the absent event in some other cases (e.g. the Bomber example). This is a problem: there is no consistent way of locating the omission which will avoid the non-locality problem.

6.1.3 Proliferation

A separate problem compounds the issue. It is surely not metaphysically salient that we expected the gardener to water the plants (or that they had promised to, or were paid to, or usually did...). Metaphysically speaking, the failure of the gardener to water the plants is on a par with the failure of the next door neighbour to water them, or a failure of aliens to water them or even a failure of the Queen to water them. The death of the plants counterfactually depends on every one of these absences since, had any one of them not been absent, the plants would not have died. For every way there could have been a watering, there is a failure of that way to occur. Not only does this give us vastly more causes of the plant's death than we might ordinarily assume, but it also compounds the issue of location. Every problem we had in locating the gardener's failure is now multiplied by every way in which the failure to water might not have obtained—by the neighbour, aliens, Queen and so on.

Perhaps there is a convergence argument we can mount to establish the location of failure. All of the ways in which the omission could have failed to obtain share one region in common, i.e. the region around the plants. The options for where to locate the failure converge on a sort of 'shell' around the plants.[2] We could take the event to

[1]This *standard* prevention case has Fighter shoot down Bomber to prevent the attack on Target, whereas a *double* prevention case has some further event, say an under-fueling, which prevents Fighter from preventing Bomber. This sort of more complicated case drives home the non-locality point yet more emphatically and poses a worry for those who wish to endorse that absence causation is genuine and that causation is a largely intrinsic matter (see Hall [2000, p.201–202]; [2004b] and Lewis [2004a, p.84]. The solution I will go on to give is not impacted by this embellishment and so I will stick to the standard cases of prevention in this chapter for simplicity.

[2]Reference to shell in this context I take from Frisch [2010]

occur there. Further we could treat the event there as essentially involving the failure to water the plants such that it has counterparts in worlds where there is no watering but not in worlds where someone (anyone) waters the plants. Thus all of the closest worlds where the failure does not occur are worlds where someone waters the plants and so are worlds in which the plants do not die. So far so good. Even better, worlds in which the gardener waters the plants will be far closer than those where it is aliens or the Queen who do and so this treatment also explains why we prefer to say that it was the gardener's failure rather than that of the aliens or the Queen.

Unfortunately this strategy will not work. The gardener is a mile away napping on a couch whilst the neighbour is just next door, so a far smaller 'miracle' is required to have the neighbour water the plants than to have the gardener water the plants, so the neighbour-watering worlds are closer still than the gardener-watering worlds. This means that the closeness-of-worlds solution cannot explain why we prefer the gardener to the neighbour as a cause of the plant's death. Lewis proposed that we take the contextual cues—that we had an agreement with the gardener, that she had taken payment, that she normally did the work—to indicate the greater salience of the gardener, as opposed to the Queen, the aliens or the neighbour. This proposal accepts the truth of the claims that the Queen, aliens and neighbour all caused the plants death[3] by their failure, but simply elevates the gardener to the top of the candidate list on the basis of relevance.

This pragmatic solution bites the bullet of the problem of proliferation but explains away our qualms about the proliferation and oddity of true negative causal claims. However, McGrath [2005] argues that Lewis's explanation is not a satisfying one. If Lewis simply had to explain why we do not assert irrelevant truths such as 'the Queen's failure to water the plants caused them to die' then an appeal to a Gricean pragmatics [Grice, 1968] may indeed offer justification. However, McGrath argues, the claims about the Queen and the alien are not merely irrelevant but are outright false, as evidenced by the fact that we will assert their negation: 'the Queen's failure to water the plants did not cause them to die'. Grice offered an account of why we might not assert irrelevant truths but, McGrath contends, he did not offer an account of why we would explicitly deny those truths. An appeal to Gricean pragmatics does not do the work that Lewis requires.

An alternative solution, offered in different forms by McGrath and Menzies, treats absence causal claims (and possibly causal claims in general) as identifying a deviation from a norm. The anticipated norm (be it a norm of regularity, etiquette, morality or any other sort) has it that the gardener should have watered the plants. Thus, the failure to water the plants is the failure of the gardener. There was no equivalent norm in place for the Queen or the aliens. Even if there was some such norm in place for the neighbour—it would be kind of them to water an obviously drooping plant—this is outweighed by the normative burden on the gardener who promised to, accepted payment to and regularly did water the plants.

[3]Menzies, personal correspondence, points out that this argument hinges on assuming a naturalistic similarity criterion for worlds such that the neighbour-watering worlds are closer. If, on the other hand, promises, duties and such like also factor in the ordering of the worlds, this result does not follow. I largely agree and I will discuss such normative options later in the chapter.

Note that this solution must be more than simply a way of understanding the pragmatics at work in the context, otherwise the 'assert the negation' response will re-emerge. If a negative causal claim is to be literally true then the normative dimension needs to be built in to the meaning of the word 'cause' in order to make it *true* that the gardener's failure caused the plants to die but literally *false* that the Queen's failure did. This suggestion has the merit of avoiding a proliferation of causes and identifying the gardener above the Queen or the neighbour as the cause of the plants' death. However, treating causation as a normative notion does not, by itself, resolve the issues of location and non-locality. What is more, it has one serious consequence: if causation itself, and not just some pragmatic selection on our causal talk, is normative, it no longer conforms to what Menzies identified as our 'central causal platitude': that causation is a natural relation. If what is or is not a cause is subject to mind-dependent notions such as expectation or etiquette then it is no longer a mind-independent matter.

Taking stock: It seems that if we treat absence causal claims as genuine then we are going to have to accept that they establish non-locality on the cheap as there is no consistent option for locating the absences that avoids action-at-a-distance in fairly ordinary cases. We may be able to respect our strong preference for identifying the gardener and not the Queen or neighbour as the cause of the plants' death but only at the expense of the assumption that causation is a natural relation. A satisfying solution should locate our absences, avoid establishing action-at-a-distance and respect the seemingly normative selection that we apply to our negative causal attributions without giving up on causation as a natural relation. In the next section I look at the existing proposals from those who accept that there is genuine absence causation in our world (Lewis, Menzies, Schaffer) and from those who do not (Beebee, Dowe) and then go on to offer my own proposal, consistent with my earlier commitments.

6.2 Existing Approaches

6.2.1 Accepting Absence Cases as Causal

Lewis's [2004a] approach to the problem of profligate causes was to cite pragmatic parameters which would dictate which true causal claims were salient and which were not. On this view, pragmatic concerns justify our preference for asserting that it was the gardener rather than the queen who caused the plants to die—in the relevant context we simply expect the gardener to have done it and so the failure to water the plants is a failure of the gardener in particular. This selection highlights one of the many true causal claims that could have been made and elevates it on account of its salience in the context, just as it is true that the big bang is the cause of every subsequent event, and yet it is rarely salient to mention it. However, as it treats absence causation as genuine, this solution falls foul of the problems of location and non-locality. It accepts profligate causes but fails to satisfactorily account for McGrath's datum that we will assert the negation of what Lewis takes to be true but irrelevant claims.

Contrastivists such as Menzies and Schaffer offer semantic accounts and argue that for any two-place causal claim 'c caused e' we must read the context in order that

we semantically complete the claim. A semantically complete causal claim, they say, requires a four-place relation involving c, e, c^* and e^* (where c^* and e^* are specific alternatives to c and e respectively) so once we have the c and e of the binary assertion, we must further read the context to work out which c^* and e^* complete the claim. Thus it is context which provides us with the salient alternative to c in the un-watered plants case: c^* is read as 'the gardener watering the plants'. This being the case, the contrastivist interprets 'the gardener's failure to water the plants caused them to die' as 'the gardener's failure to water the plants, rather than water them, caused the plants to die, rather than not die'. Further, the contrastivist can (and Schaffer does [2005, p.301]) take the negative nominal to pick out an actual event, such as the gardener's nap. We would perhaps be squeamish when it came to asserting that the gardener's nap caused the plants to die in a binary mode, but that simply explains why we use the negative nominal in order to trigger the correct contrastive. Holding the contrasts fixed we can substitute-in the positive nominal for the same event and the contrastive causal claim still rings true: 'the gardener's having a nap, rather than watering the plants, caused the plants to die, rather than not die'.

Perhaps the contrastivist can locate the negative event, but it is not obvious that they can stifle the proliferation of causes. Surely it is true that had the Queen watered the plants they would not have died. If so then the contrastive claim 'the Queen's failure to water the plants, rather than her watering them, caused them to die instead of not die' is true. This can be iterated for the neighbour, the aliens or the velociraptor that didn't water the plants either and so creates just as great a proliferation of causes as a non-contrastive account. Of course, the contrastivist can say that this is not a relevant claim to make, but McGrath's point against the pragmatist re-emerges: it is not just infelicitous to say that it was the Queen's failure to water the plants that caused them to die, many would assert the negation of this and so it must be outright false and not just irrelevant. The contrastivist needs to explain why the binary claim is false and so must argue that the binary claim does not yield the true quaternary claim that I offer above. Schaffer argues [2005, p.354 fn.10] that the positive event that we pick out when we talk of the Queen's failure to water the plants is in fact some regal event, such as a feast, that she is occupied with at the time. The salient alternative to such a feast is some other queenly event, not watering the plants, so the correct interpretation of the binary claim that the Queen's failure to water the plants caused them to die is: the Queen's attending a feast, rather than attending to her corgies, caused the plants to die rather than not die. This claim is false and explains our willingness to assert the negation of the binary claim, according to Schaffer.

I think this rides roughshod over the binary claim which was very much about the Queen, the plants, and water but was interpreted as being about a feast and corgies instead. You need know nothing of what the Queen was doing, or that she has corgies, in order to reject the claim that her failure to water the plants caused them to die. I just do not think Schaffer's interpretation is plausible. Menzies adopts a different strategy, one which takes the contrast to be supplied by normative considerations. In relation to a parallel example he says:

> [T]he contrast between the doctor's omission and the normal course of

> events explains why the patient died rather than survived, but there is no comparable explanatory contrast between the hospital cleaner's omission and the normal course of events. [Menzies, 2009, p.364]

Analogously, the gardener has promised to water the plants, taken payment for the service and has done so consistently in the past. Relative to a range of different norms (moral, contractual, regularity) the gardener can be expected to water the plants but there is no norm at play in which the Queen can be expected to water them. Citing the Queen's failure to provide water does not explain why the plants died, because the Queen had nothing to do with the violation of the relevant norms (moral, contractual and behavioural). To truly explain the plant's death we need to explain why the expected failed to materialise and talking about the Queen is simply not an explanation.

Whilst this suggestion is an improvement on Schaffer's, I note three deficits: First, by reifying omissions the problem of non-locality emerges. Either the omission is located where the positive event is in the actual world, i.e. at the gardener, in which case it causes at a distance, or it is located where the would-be event would have taken place, in which case it acts at a distance in prevention cases, such as that of Bomber and Target. Second, Menzies solution requires that absence causation is itself normative in nature. In his own words this 'violates the strictures of causal naturalism' [2009, p.364]. Third, Menzies shifts the emphasis from causal truth to explanatory virtue. It is the explanatory force that makes one causal claim acceptable and another not, not truth. If we were analysing causal explanation, then perhaps this approach would be reasonable—in fact I think that it is what the contrastive project does analyse and I think that, in that context, it is a reasonable approach—but I am interested in the causal relation itself, not just causal explanation. Menzies appears to have shifted the topic.

6.2.2 Rejecting Absence Talk as Causal

The pragmatic and semantic camps, as I characterised them in the last section, are both accepting of our negative causal attributions and both seek to include them in their causal theory. An alternative strategy is to take the metaphysical issues around absence causation to indicate that the claims themselves are problematic. A simple error-theory of absence talk will not do as our negative causal attributions are simply too pervasive—perhaps our most pressing need to understand the causal workings of our world are cases of death, but all cases of death are cases of the absence of life via the absence of blood, or oxygen or whatever else we need. So, if absences are to be dismissed as non-causal, our negative causal claims need to be explained, not just written off as aberrations. Here I discuss two such explanations.

Dowe [2001] has argued persuasively that our most problematic negative causal claims, including simple absence cases like the un-watered plants and more complex cases of prevention and double-prevention, can all be understood as would-be causal claims rather than actual causal claims. The idea is that when we say that the gardener's failure to water the plants caused them to die, we are really saying that if the gardener had watered the plants then they *would* not have died. As such, the claim

about the failure to water is not genuinely a claim about our world but instead a claim about some other world in which the gardener does water the plants. This will often be useful to cite, and does concern causation of a sort, just not actual-world causation, and so Dowe considers such cases examples of what he calls 'quasi-causation' rather than actual-causation. This move allows Dowe to avoid tricky questions about the actual-world location of the negative events since, on his view, there is no actual-world event to locate. Further, the proliferation of genuine causes does not occur on this view as only actual-world events are genuinely causal and it is merely the would-be quasi-causes that proliferate.

One might complain, as Schaffer has [2004a], that our negative causal ascriptions carry more weight in our everyday lives than such a second class status would allow. We take tremendous pains to avoid harm, prevent damage and stop erosion and we do so as part of otherwise positive causal paths: we put on a protective suit, catch a ball or apply a varnish to cause these negative events and then we go on to swim, to throw and to sail as a result of having done so. Absences slot neatly into causal chains that are canonical examples of causation, so relegating them to second-class 'quasi' status, or indeed insisting that they have a fundamentally different status, requires significant motivation.

It is worth noting that there is broad disagreement in the literature about the status of absence causation.[4] Such disagreement may well indicate that absence causation is *some* sort of special case deserving of a distinct treatment. Note too that Dowe's revision does not involve rejecting all talk of negative causes, but instead admits that such claims do share the concept of causation, just at a remove from the actual world. Where Dowe holds that actual causation is to be analysed as a physical process (i.e. not in counterfactual dependence terms), he holds that quasi causation is the same type of physical process in some other world, some would-be world, which does not in fact obtain. Of course this sort of view is an affront to those partisans who take absence causation to be as genuine as any causation can be, but by stepping back to see the literature as a whole, we can appreciate the bi-partisan nature of Dowe's solution. The would-be causal approach accommodates the intuition that preventions and omissions do play *some sort* of causal role, but it does so without opening up the problems of location, profligacy and non-locality.

I note two issues for my adopting Dowe's approach: First, Dowe analyses causation as a physical process, not as counterfactual dependence. I will not debate the relative merits here—it is a subtle and complex issue and one that hinges on whether there are any plausible counterfactual theories of causation, which is, in part, the topic of this thesis—however I will point out that the nature of Dowe's analysis of causation, and therefore of negative causation, only works in worlds with laws like ours whilst counterfactual accounts have broader ambitions. Secondly, in the face of problems concerning the closest-world treatment of would-be causal claims Dowe [2009d] embraces a causal-modelling semantics. Again, entering the causal-modelling debate is not the objective

[4] A cross-section of the literature discussing negative causation : Beebee [2004], Bennett [1988], Bernstein [2014], Bernstein [2013], Collins [2000], Collins et al. [2004], Dowe [2001], Hall [2002], Hall & Paul [2003], Hall [2000], Lewis [2004a], Lewis [2004b], Mellor [2004], Menzies [2009], Sartorio [2010], Schaffer [2004a], Weslake [2013b].

of this thesis,[5] but I do point out that Dowe requires a significant theoretical departure to handle these cases.

Finally, Beebee [2004] has argued that our commonsense intuitions concerning negative causation, those intuitions that Schaffer, Menzies and Dowe are so keen to respect in their causal theories, are unstable. Beebee argues:

> [C]ommonsense intuitions about which absences are causes and which aren't are highly dependent on judgements that it would be highly implausible to suppose correspond to any real worldly difference at the level of the metaphysics of causation. For instance, sometimes common sense judges the *moral* status of an absence to be relevant to its causal status. But no philosopher working within the tradition I'm concerned with here thinks that the *truth* conditions for causal claims contain a moral element. It follows that whatever we think about whether or not causation is a relation, we're going to have to concede that common sense is just wrong when it takes, say, moral differences to determine causal differences. There is no genuine difference between those cases that common sense judges to be cases of causation by absence and those that it judges *not* to be cases of causation by absence. Hence... commonsense judgements about causation by absence are often mistaken. [Beebee, 2004, p.293]

Note the shift in dialectic here between Schaffer's criticism of Dowe and Beebee's argument against our intuitions about negative causation. Beebee is operating on the assumption that we are trying to identify some mind-independent, real-world relation, what Strawson has called a natural relation, and so intuitions that are inconsistent with such a relation are to be questioned. Schaffer, by contrast, takes the intuitive and conceptual role played by absences as base data and so any theory, or naturalist assumption, that conflicts with the data has failed to live up to its billing as a theory of that concept.

An explanation is due from those who wish to undermine the negative causal data as to why negative causal intuitions are formed, why they are so persuasive and why they are so prevalent. Dowe offers quasi-causation as a way of understanding the negative causal claims but Beebee makes the less-conciliatory claim that there is no negative causation. Negative causal claims, and the intuitions which fuel them, are to be understood not as literal causal claims but instead as causal *explanations*. Following Lewis [1986a, p.217], Beebee advocates understanding causal explanations as aiming to "provide some information about [the] causal history" of an event. Importantly, this approach allows one to give negative information about the causal history without that negative information being *part* of the causal history. By analogy: saying "Neil is not nine feet tall" provides information about my height without actually stating my height. Driving this wedge between causation and causal explanation allows Beebee to maintain that there is no genuine negative causation, but that negative causal claims can still make sense if they are taken to be causal explanations rather than direct causal claims. This is revisionary in the sense that it requires a re-interpretation of the literal

[5] Though I will compare my ACCT Analysis with causal modelling accounts in Chapter 9.

form of what was said, but it is a modest revision when compared to the contrastivist's four-place reading of our binary causal claims.

Beebee's solution also fares well in relation to the problems of locating absences, and of profligate causation. Absences are not events and so claims concerning absences do not need be spatiotemporally located. This means that there is no problem of locating the cause in a negative causal claim as there is none to locate. Also, it is explanatory to cite the failure of the gardener to water the plants, but it is not explanatory to cite the failure of the Queen to water them. This is what makes the gardener, and not the Queen, the right absence to cite.

I note three issues with Beebee's view. Sometimes we really *do* want to locate an absence in a particular place—such as the absence of oxygen in the lungs, or absence of food in the stomach—so saying that absences never have location is at least counterintuitive. Further, we offer something more than mere correlation when we identify the lack of water as being relevant to the death of the plants and we would expect a good account of negative causation to explain that. As it stands, saying that there was no water, and that the plant died succeeds in being an explanation simply in virtue of providing some information about the causal history. However, negative causal claims do not simply state that there was no water and that the plants died, but rather that there is some meaningful connection between the two occurrences such that the plants died *because* there was no water. So, just giving some information about the causal history is not enough, we need to give the right sort of information. Finally, McGrath [2005] has complained that Beebee's approach cannot account for our literal denial of those outlandish negative causal claims involving the Queen, nerve gas or a velociraptor. According to Beebee's view each such claim ought to be read as an explanation, and each is a true explanation of the effect in question. Why, then, do we consider such assertions to be literally false?

6.3 A Positive Proposal

In this final section of the chapter I will introduce my own proposal. First I will recapitulate my commitments so far, then endorse a normative, explanatory, reading of our negative causal claims. I will then argue that this view meets the desiderata outlined at the end of 6.1 and conforms with the counterfactual view defended so far in this thesis (dubbed the ACCT Analysis). Finally I will show how the proposed treatment of absence causation relates to a recent proposal from Weslake.

6.3.1 The ACCT Analysis

Recall my ACCT Analysis from Chapter 5: actual event e causally depends on actual event c iff c and e are distinct events and the following conditional is true: $\neg Pc \; \Box\!\!\rightarrow \neg Oe$; c is a cause of e *simpliciter* iff c_i and e_i are connected by a chain of causal counterfactual dependence; a causal assertion of the form 'c caused e' is true in context C iff $e_n$is connected by a chain of counterfactual dependence to both c_n and c_i, where counterpart relation n is a function of the context C.

So, for any events c and e there are three tests to establish the causal connection between them:

1: ACCT Canonical Test $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$ (or chains thereof)

If this conditional (or chains thereof) is true for c and e, then c is a cause of e simpliciter.

2: ACCT Actual Test $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_n$ (or chains thereof)

If this conditional is true for c and e then c made some *specified* difference to e—i.e. the difference concerning the essential features in that context.

3: ACCT Contextual Test $\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n$ (or chains thereof)

This test establishes that a *specified* difference in c makes the specified difference to e—i.e. the differences concerning the essential features of c and e in that context.

Applied to the case of the gardener, not all the conditions are met. As a quirk of our physics, the gardener napping on the couch does make some difference to the plants as she exerts some gravitational pull on them, so the Canonical test (1) is met. The failure-to-water aspect of whatever else the gardener is doing (i.e. napping on the couch) does indeed make a difference to the dying of the plants for had she not had that aspect (i.e. had she watered the plants) they would not have died. This means that the Contextual test (3) is also met. However, the Actual test (2) is not met. The clean excision of c does have some impact on e thanks to our physics, but that clean excision does not make a difference to the aspect of e that we are interested in—the dying of the plants. Thus, on the ACCT Analysis, 'the gardener's failure to water the plants caused them to die' is false.

So it also goes in the case of Prevention. The event in which Bomber is downed does exert a (merely gravitational) influence on Target and, had Fighter not shot down Bomber, Target would have been destroyed so tests 1 and 3 are satisfied. But the clean excision of the event, that is the total removal of that event including Fighter, the missiles and Bomber and its wreckage, without any replacement with something similar, does not make the difference to Target's survival—it survives anyway. Thus, on the ACCT Analysis, 'Fighter's shooting down Bomber caused Target's survival' is false.

So, on the ACCT Analysis I advocate, certain negative causal claims will turn out straightforwardly false. Notice this is not a stance motivated by the problems of location, non-locality or profligacy, but rather a direct consequence of the ACCT Analysis already on offer. This means I owe an explanation of why negative causal claims are so prevalent and why they seem so natural.

I will shortly flesh out my proposed account of absence causal claims but before I want to highlight two features of the ACCT Analysis. First, absence cases fail the three ACCT tests in a distinctive way—they pass the third test, but not the second.[6]

[6]That they pass the first in the examples given is a quirk of our physics in which minor gravitational influences are exerted by every region with mass. I do not rest my case on this contingent feature.

I will refer to these cases as F-T cases in reference to the False value for the second conditional and the True value for the third. The F-T structure is only possible where one aspect of the c event correlates with one aspect of the e event, but where there is no intrinsic feature of the c event which alters the salient aspect of the e event. If correct, this allow us to identify absence cases purely by their counterfactual structure.

Second, not all cases involving a negative nominal will fail the test. Those situations where a positive state of affairs within a region is merely picked out by a negative nominal, such as where my lack of technique when shooting is relevant to the wayward result, will pass it: to cleanly excise the arrangement that constituted a lack of technique is to cleanly excise my shot altogether, not to replace it with a graceful version. So excised, there is no result which can be wayward. Such cases are not genuine cases of absence causation but merely positive states re-described. The proposed test tells the two apart in a way that I think is in line with Lewis's observation regarding causation by omission:

> It is one thing to suppose away the event *simpliciter*, another thing to suppose it away *qua* omission. [Lewis, 1986d, p192-193]

My test requires that both be satisfied for genuine causation.

6.3.2 Would-be Causation

Those who think that there are genuine cases of absence causation in the actual world (Lewis, Menzies and Schaffer are such) will surely complain that the ACCT Analysis has gotten the wrong answer and should be rejected on those grounds. However, those who deny that there are genuine cases of absence in the actual world, such as Dowe and Beebee, may see this as a positive result. For my part I take seriously the intuitive appeal of absence causal claims and I take it that I owe an explanation as to why they are so appealing if they are in fact strictly false.

I think the first step is to understand that absence causal claims are not just claims about some absence or other, they refer to the absence of some specific alternative. To suppose away the absence of that alternative is just to suppose that the alternative did in fact occur. Unlike causal claims about throwings, floods and explosions, causal claims about failures-to or lacks-of are focussed not on the actual world and its occurrences but rather on another world in which things go differently—a specific sort of differently.

If absence causal claims have this implicit other-worldly focus, then it seems that they are claims about some other world which reflect the state of the actual world by contrast with the alternative—the actual world is just one of many ways in which the specific alternative failed to occur. This suggests that the absence claims are centred on some other world, not ours, just as Dowe has argued. When we talk of a world where Oswald doesn't shoot, and where Kennedy doesn't die in Dallas, we characterise that world by pointing out what it lacks relative to ours. I suggest that when we talk of omissions we conduct a sort of Copernican shift: our world is no longer the centre, some other world is, and it is relative to that world that we characterise the events in our own.

This view explains our problem of location. The location of counterpart events is often indeterminate—where do we locate candidate counterparts of the fight between Ali and Foreman in worlds where the two men never meet? Do we locate it with Ali? With Foreman? Or in a boxing-ring shaped space in Kinshasa? Such ambiguity in finding the counterparts exists even when there is no ambiguity about the location of the target event. If I am right, and absence causal claims are centred on some other world, then our world contains merely the *counterparts* of the target events. This would explain why we are ambivalent about the location of the failure to water the plants in the actual world since there is some determinately located watering in another world, our world only contains counterparts to that watering and, as we have just seen, the location of counterpart events is often ambiguous.

Borrowing heavily from Dowe, my proposal is not simply that the target events of our absence talk are other-worldly, but that the causal relation between them is too. I take the claim 'the failure of the gardener to water the plants caused them to die' as being acceptable because 'the gardener's watering the plants caused them to survive' is true in the alternative world. In Dowe's early discussion of this *would-be* causal strategy [2001] he specifies that the solution is supposed to allow you to plug in your own ("B.Y.O." [p.221]) semantics for the other-worldly causal claim. That means that the causal claim in the other world is subject to just the same causal analysis as I offer in the actual world, *modulo* which world acts as the relevant centre, i.e. which world contains the target events, when considering which worlds are *closest*.

This view explains our problem of non-locality. The events that are causally related are not those in the actual world, which are merely counterparts, the events that are causally related are those in the alternative world. Sufficiently close alternative worlds will share our physics and so we can expect most of the causal connections to happen as ours do: by a chain of spatiotemporally continuous occurrences. It was a mistake to think that in absence cases the ordinary, proximal, causal connections in the alternative world meant that their counterparts were causally related in ours. It was this mistake that gave rise to the problem of non-locality.

This strategy succeeds, then, in two of the problems raised in part 1. As it stands, however, the strategy does not address the problem of profligacy and fails to specify how we are to pick out the all-important alternative world. This is related to the problem of nested counterfactuals that Dowe [2009d] later raises for his initial "B.Y.O." semantics proposal from [2001]. In the next section I will explore this issue and suggest a normative strategy which locates the alternative world and addresses the problem of profligate causation.

6.3.3 Problems with Would-be Causal Semantics

In his [2009d] Dowe argues that the Lewisian semantics won't work for negative causal claims and so it is not B.Y.O. semantics after all [p.703]. I will consider the two problems he raises here before positing my own solution.

The first problem stems from the closeness of the closest *would-be* worlds. The closest worlds where you *suppose away* something that did occur can be problematic enough (see [Lewis, 1979]) but the worlds where you *suppose in* something that did not

are worse because they require a greater miracle to bring them about. To suppose that the rock had not been thrown at the window we can enact a relatively minor 'miracle' upon Suzy to change her mind about throwing and so ensuring that she holds onto the rock or drops it instead. To suppose that the plants had been watered, however, requires that we enact a far larger miracle, bringing the gardener and water to the right spot at the right time. But where does the miracle begin? A small miracle in the past may have been enough to bring the gardener to the plants in an otherwise unremarkable way, but if we hold fixed the past until shortly before the would-be watering, then the miracle required to bring the gardener suddenly from her sofa several miles away is a very large deviation indeed. Lewis's standard methodology is to hold fixed the past as much as possible, but not at the expense of large-scale miracles (again, see Lewis [1979] for details), so it seems that we should favour the small miracle in the past over the large-scale miracle. The problem with this is much the same as the problem of backtracking counterfactuals in general, according to Dowe: the altered path the gardener takes to water the plants in the small-miracle scenario means that she cannot go home, feed the cat and go for a nap as she does in the actual world. This means that, in this alternate world, her watering the plants prevents the cat from being fed. This sounds odd enough but becomes particularly bad when we consider that the watering of the plants happens at a later time than the cat's being fed—Hall was worried about non-locality on the cheap but here we seem to have backwards causation on the cheap!

Of course I do not think that we do have backwards causation on the cheap here at all, as a little care will reveal. If we are considering the consequences of an earlier deviation then the causal supposition is not simply: had the gardener watered the plants, the cat would not have been fed, but rather: had the gardener come to work and watered the plants *as she was supposed to*, then the cat would not have been fed. The difference is subtle but in the second the antecedent alludes to the path that took the gardener to the plants, not merely the act of watering in isolation. On this interpretation the failure to water the plants is not just a localised event around the plants, but rather a whole alternate path that would have occurred had history diverged at some earlier point. That alternate path has many consequences but that should not be confused with the final step on that path—the watering—having those consequences. Had she made the decision to go and water the plants, that *decision* would have prevented the cat from being fed (perhaps that is why the gardener deviated in the first place!). This decision is prior to the feeding and so does not establish backwards causation on the cheap.

Dowe's second complaint is more serious. First, Dowe's analysis of would-be causation takes would-be causal claims to have the following counterfactual form $c \mathrel{\Box\!\!\rightarrow} [c$ causes $e]$ where this is to be understood as: in all of the closest c-worlds it is true that c causes e, and where $[c$ causes $e]$ is understood by whatever semantics of causation turns out to be the right one. On the assumption of a simple counterfactual account of causation[7] this becomes $c \mathrel{\Box\!\!\rightarrow} [\neg Pc \mathrel{\Box\!\!\rightarrow} \neg Oe]$ which is to be read as: in all of the closest c-worlds it is true that had c had not occurred then e would not have occurred.

[7]This example does not require the resources of the three-step test I propose and so, for simplicity, I demonstrate the point with simpler apparatus.

For this claim to be true all of the closest $\neg c$-worlds to all of the closest c-worlds must be $\neg e$-worlds too. Note that when we suppose-in the closest c-worlds we perform a law-defying miracle, and from each of these various c-worlds, with each of their deviant laws, we must perform a second law-defying miracle from that world and now check to see if all these two-step removed $\neg c$-worlds are e-worlds.

Following Barker [1999] and Jago and Barker [2012], Dowe does not believe we can be confident in such assertions. Here is an illustrative example adapted from Barker [p.430]: Fred was booked to travel on the ship but cancelled at the last moment. A week later he reads that the ship has sunk with the loss of all lives and thinks 'had I gone, it would have caused my death'. On Dowe's reading, this is only true if in all of the closest worlds where Fred boards the ship it still goes on to sink, he dies and *had he not boarded the ship he would not have died.* This last clause is important so that we get causation between the supposed-in boarding and the death, not just correlation. However, this clause requires that we assess a two-stage counterfactual: the first stage violates some laws by supposing-in Fred's boarding; the second stage violates some more by considering the counterfactual *from the perspective of the supposed-in worlds.* So, to assess the embedded counterfactual we are considering what happens at worlds at a two-miracle remove from the actual world. How can we be sure that all of those primary worlds are worlds where Fred dies *and* be certain that all of the secondary worlds are worlds where he does not? We do not have to be certain of this example to see that the truth or falsity of such a claim is very difficult to assess. We are much surer of our absence claims, and of the claim that had Fred boarded the boat he would indeed have died. So it seems that the embedded counterfactual account is a poor analysis of the negative causal claims.

I think that we can take a lesson from these objections, but it is not the lesson that Dowe himself takes. Recall that Dowe is a process theorist about causation and does not hold to a counterfactual analysis of causation. Dowe's conclusion is that the problems with these cases stem from plugging in a counterfactual analysis of causation. However, his own preferred *probabilistic* account of causation is in trouble if it is plugged in to counterfactual structure too. So long as the boat does not sink in at least one of the closest supposed-in worlds, an outcome that is likely (as Barker [1999, p.430] points out because boats rarely sink), then the claim that Fred would have died comes out false, just as it does when a counterfactual test is plugged in instead. This strongly suggests that the problem is with the would-cause semantics, not the causal semantics which are used. So, the lesson I take from his objections is that we need another way of specifying the alternative world(s) that our absence causal claims imply. In the next section I am to answer that need with a semantics of would-be causal claims that is consistent with a range of causal accounts—it will be B.Y.O. semantics once again.

6.3.4 A Norm-centred Approach

I propose that we view the alternative world in which the would-be events occur as a *normal* world, in some very broad sense of normal. This world is not necessarily the closest world where the supposed in event occurs, but is instead the closest of the most *normal* worlds where the event occurs. I follow McGrath and Menzies in my

use of *normal* here, as meaning almost any kind of norm: regularity-based, moral, social, contractual, legal or whatever else and is supposed to capture the framework of expectations that the person making the claim brings with them when they suppose in the event. So, the absence of heating causes the room to be cold by the lights of the person expecting heaters, but not by the lights of the person who doesn't. The first models a normal world in which there are heaters, which stop the cold in that world, but the second models a different normal world in which there are no heaters and so they never supposed in the antecedent condition.

So a would-be causal claim is analysed in whatever way a causal claim in the actual world is, but the claim is taken to refer to some close normal world, not the actual world or closest possible world. The causal claims can be thought of as centred on the normal world and so I call this approach a 'norm-centred' account of would-cause semantics. Which normal world such claims are to be centred on can be highly dependent on the expectations of the person making the claim but where the norms are statistical or are expectations supported by our natural sciences, then that subjective dimension will not necessarily undermine general agreement on which would-be causal claims are acceptable. Those who do disagree on what the would-be scenario entails can be expected to disagree on the causal attributions in that would-be scenario. None of this changes the causal facts in the actual world, or the causal facts at any given world, it simply varies *which* alternate world is being considered. So, on this view, it remains plausible that causation is a natural relation between events despite our negative causal claims having a substantial normative dimension.

I think this approach neatly resolves Dowe's first worry about the miracles required in would-be causal claims when plugging in a counterfactual analysis of causation. The gardener was *supposed* to come to work and water the plants and so that is what we model when we consider the closest worlds in which 'the gardener watered the plants' is true. Perhaps they are not *the* closest worlds that get the watering to occur (the neighbour does it in those), but they are the closest worlds in which the normal trajectory is maintained and hence it is the normal worlds, not the closest ones that our would-be causal claims should be centred on.

It also resolves the second worry about nested counterfactuals. A norm-centred view does not look to all of the closest worlds where Fred gets on the boat but rather some closest normal world in which he does. The normality of that world shifts when he finds out about the sinking of the boat. Before reading the news he supposes that had he gotten on the boat he would have gotten to the destination. After the news he updates his assessment of the closest normal worlds and he now supposes that had he gotten on the boat he would have died. The shift concerns what Fred holds fixed when he models the alternative world and that is sensitive to things like what he knows and what he expects. When supposing in Fred's getting on the boat, if we simply hold the past fixed up until the point that he got on the boat then the boat may well sink in one of the (very many) worlds in which he embarks, or the causal counterfactual 'had Fred not boarded he wouldn't have died' might be false at that world. That is why the nested-counterfactual strategy fails. However holding fixed many elements of the past and some specific element of the future—the sinking—tracks what Fred was really meaning: had he gotten onto the boat, and if it had gone on to sink as it did

in the actual world, then his getting on the boat would have caused him to die. By taking the alternative world to be normal in respect of Fred's updated expectations, the norm-centred view tracks Fred's assertions before and after the news report about the sinking.

Perhaps most importantly, the norm-centred approach helps us get a grip on the problem of profligate causes. Recall that the absence of watering by the gardener was on a metaphysical par with the absence of the Queen, aliens or the next door neighbour doing the watering instead. Worlds in which the gardener does water the plants are closer than worlds where the Queen or some aliens do and this lends support to a closest-worlds approach to supposing in the watering since it tracks our preference for citing the absence of the gardener, over that of the Queen or aliens, for her role in the death of the plants. The closest-worlds approach does not, however, track our preference for citing the absence of the gardener over the absence of the neighbour. If it takes a smaller miracle to suppose in the neighbour watering the plants, then it is the neighbour and not the gardener who we should, according to the closest-worlds approach, cite as the cause of the death of the plants. And yet we do not.

The norm-centred view on the other hand does track our preference for the gardener over all other candidates, precisely because it models the world we would have expected—the one with the gardener doing her job, what she promised to etc. The person who thinks that the Queen is the most normal candidate to water the plants (perhaps they think she promised to or had a secret agreement with the gardener) will deny that it was the gardener's failure to water the plants that caused them to die, in just the same way that most people will deny that the plants died because of some omission by the Queen. If correct this would show that our expectations determine which absence claims we are willing to make and which we are not. In fact, this seems to explain McGrath's datum that we will assert the negation of those more obscure absence claims since a watering by the Queen, aliens or velociraptor simply do not feature in the world that we are modelling. They are not present to do the causal work in our would-be scenario and so it is literally false to cite them as a cause at *that* world.

Allowing this degree of normativity into the treatment of absence claims is familiar from the work of McGrath and Menzies and so a familiar objection can also be raised. If you are analysing causation in terms of normativity then you have just given up the central causal platitude that causation is a natural relation. As I have shown above, however, the normativity concerns which would-be world is under consideration, not which causal claims are true at that world. Causation remains a natural relation regardless of whether there is a normative influence upon would-be or possible causal claims.

6.3.5 Explanation

I have indicated which causal claims are supposed to receive a would-be reading—those which have a F-T profile on the Actual and Contextual ACCT tests—and I have proposed an account of the would-cause semantics that resolves the existing problems. What remains is to clarify the resultant status of our absence causal claims: why do we say that the absence of one thing caused another if that is, strictly speaking, false?

Here I borrow heavily from Beebee. Absence causal claims are literally false in our world but their positive correlates may or may not be true in some other close normal world. For example, 'the absence of rain caused the reservoir to be empty' is false in our world (since it fails the ACCT Actual test), but in some close normal world where the rain did occur, the reservoir is full. From the vantage of the close normal world, the rain caused the reservoir to be full, as had there been no rain, there would not have been a full reservoir. However absence causal claims made in our world do give some information about the causal history of our world, albeit negative information about what didn't happen, and, following Lewis, that makes them candidates to be considered causal explanations of what occurred or failed to in our world. Absence causal claims are not causal claims at all, they are causal explanations.

Nontheless, they are not always good explanations. Where our expectations about the would-be scenario differ, such as when you expect the Queen to do the watering and I do not, then the explanation that I provide for the death of the plants (that the gardener failed to do it) will rank as a bad explanation in your estimations. In forming each rival explanation we have no need to address what occurred in the actual world beyond verifying that neither the Queen nor the gardener watered the plants and that the plants did indeed die. Any world where we suppose-in a waterer will save the plants but our strong preference for the gardener in the standard example suggests that that not just any waterer will do. The absence of the gardener and the absence of the Queen may be on a metaphysical par but they are not on an explanatory par. Seen as explanations, our preference for one negative causal statement over another makes perfect sense.

McGrath might rejoin here that I have only accounted for our *preference* for one causal claim over another but that I have not yet accounted for her datum that we assert the negation of claims which meet the requirements for a causal explanation. Furthermore, I have been holding alternative views to account using this datum, so I cannot very well abandon it now. If McGrath's point stands against Beebee's causal explanation account, then shouldn't that same point apply against the account I have just given too?

I do not think McGrath's point stands against my account, however. Characterising Beebee's strategy, she says:

> The idea seems to be that [the omission] explains e iff the fact that [the omission] occurred (together, perhaps, with the fact that e depends on [the omission]) rules out some hypothesis about the causal history of e.

This is not how I read Beebee's account and it most definitely is not the account that I will endorse[8]. If a statement concerning an omission does convey some information about the causal history of e, then the fact that the omission occurs does formally qualify as an explanation. However 'explains' is a success term, it implies not just that any old explanation has been given, but that an adequate one has. This leaves plenty of scope for omissions which meet the formal requirement to be an explanation without

[8]Of course if I am wrong about Beebee's view in this instance, then my view is just a little more novel.

in fact successfully explaining the phenomenon under consideration. This is true for positive claims too: it may be the case that the big bang is present in the causal history of the death of the plants but that does not mean it explains their death.[9]

It is clear then what I (and Beebee) should say to McGrath's asserting the negation examples: The omission featured in the negative causal claim being negated may have formally qualified to sit in the explanatory role vis-á-vis the effect, but it was nevertheless inadequate as an explanation of that effect. The Queen's failure (or that of the aliens or a velociraptor) are simply inadequate explanations of the plant's dying and so we will assert the negation of any claim which implies the contrary.

6.3.6 Absences and Proportionality

A recent paper by Sartorio [2010] embellishes the plant watering example by introducing the Prince of Wales as an actor in the story. By specifying what the Prince of Wales would have done instead of watering the plants, Sartorio argues that counterfactual analyses of causation are forced to radically over-count causes. In this chapter I considered the proliferation of negative causes that emerge if we endorse a counterfactual analysis of causation and accept the existence of negative events. Sartorio extends this problem by showing that those who do endorse absences as causal have a related problem: there are many minor positive events (scratching your nose, reading a paper) that occur when the watering ought to be being performed and upon which the failure to water the plants depends. This being so, the counterfactualist is forced to over-count not only negative causes but positive ones too. Sartorio advocates a proportionality constraint (akin to that found in Yablo [1992]) on causes to avoid this problematic proliferation of causes: speaking roughly "nothing with a poorer essence would have been sufficient for the effect to happen, and nothing with a richer essence was necessary for the effect to happen." [Sartorio, 2010, p.17]

Sartorio's Prince of Wales problem only applies to those theories which endorse absences as genuine causes, which mine does not, so the objection does not directly impact on my position. However, a recent response from Weslake [2013b] teases out an interesting and important distinction concerning the role of proportionality. Weslake distinguishes the *metaphysical* problem that Sartorio has posed which concerns the sufficiency of counterfactual dependence for causation, and what he calls the *psychological* problem which concern our rejection of claims such as 'the Queen's failure to water the plants caused them to die'. This is a familiar distinction from the pragmatist regarding absences (i.e. Lewis) who allows for (metaphysically) true causal claims to be unhelpful or misleading. However, Weslake points out that this distinction gives us two places to consider imposing a Yablo-esque proportionality constraint. Yablo [1992] along with List & Menzies [2010] and Sartorio [2010] take proportionality to be a metaphysical constraint on which causal claims are true but Weslake instead endorses proportionality as a constraint on what makes for a good causal explanation. The diagnosis offered

[9]Perhaps all that the formal criteria for a causal explanation establishes is that there is *some* possible scenario for which this would successfully explain the phenomena, not that it explains that phenomena in every possible scenario.

is that some theories of causation may confuse principles concerning what makes for a good explanation with what counts as a true causal claim.[10]

This exchange between Sartorio and Weslake is helpful on two fronts. Firstly, Weslake is making a similar point regarding the distinction between rejecting a causal claim and rejecting that same claim in the role of an explanation. I think this gives my view additional credence. Secondly, this distinction foreshadows my discussion in the coming chapter regarding transitivity and the role of proportionality.

6.4 Conclusion

Some causal claims involving negative nominals for events will come out true on the ACCT Analysis—those where the negative nature of the nominal is dispensable, such as with the waywayd kick example I gave in §6.3.1—but others will fail the causal test in a distinctive F-T pattern. Those that exhibit this pattern should be read as making a would-be causal claim, not an actual causal claim, where the would-be situation is modelled as some close normal alternate world in which the absence is taken to occur. I call this approach to absence causal claims *norm-centred*. The norm-centred approach is a blend of Dowe's would-cause semantics and the normative approach of McGrath and Menzies but unlike the first it can account for our normative preferences and unlike the second it can hold on to my standing assumption that causation is a natural relation. The norm-centred approach escapes the problems of location, non-locality and profligate causes and it does so in a way consistent with my standing commitments in this thesis.

[10] I suspect that the causal contrastivists, who take their lead from theories of causal explanation, make exactly this mistake when they endorse absence and would-be causal claims.

7 Transitivity and Proportion

> That causation is, necessarily, a *transitive* relation on events seems to many a bedrock datum, one of the few indisputable a priori insights that we have into the workings of the concept. [Hall, 2000, p.198]

7.1 Transitivity Problems

In ordinary cases we reason via chains of causal connections to the conclusion that the first part of the chain caused the last. For example:

> Billy broke the window, which in turn set off the alarm, so Billy caused the alarm to go off.

This pattern of reasoning appeals to the intuitive notion that causation is a transitive relation. Such reasoning requires something like the following transitivity thesis:

TRANSITIVITY If c is a cause of d and d is a cause of e, then c is a cause of e.

Counterfactual dependence is not transitive and so, in order to respect the intuitive notion that causation is transitive, Lewis built transitivity into his original account of counterfactual causation by fiat. This addendum is essential for the success of such a counterfactual analysis, as is made clear in cases of early pre-emption. Recall:

> **EP:** Billy and Suzy are out to vandalise. Suzy reaches for the only rock and throws it at the window. Had she not thrown it Billy would have, and he is notoriously accurate. The rock strikes the window and the window breaks.

Here the window breaking does not depend on Suzy's throw as Billy would have brought it to be anyway. Thus, counterfactual dependence is not necessary for causation, but rather chains of counterfactual dependence are. When c causes d and d causes e, c, then d and e form a causal chain. There is such a chain that connects Suzy and the window, but not Billy and the window, and so Suzy is a cause of the window break and Billy is not. The pivotal difference is that had the rock not been at some midpoint in its flight (call this event d) the window would not have broken and had Suzy not thrown the rock it would not have reached that midpoint. Thus d is a pivotal, or *partisan*, midpoint that separates Billy and Suzy in terms of their causal role. Without the stipulation of transitivity the counterfactual analyst cannot identify Suzy as the cause—that would be a failure of the counterfactual analysis. So, a Lewis-style counterfactual analysis requires that causation be transitive and my ACCT Analysis is no different. I too appeal to *chains* of causal dependence for just the reason that Lewis does. Chains of counterfactual dependence between events makes for causation between those events, so TRANSITIVITY is built into the fabric of my account.

Whilst transitivity seems to be a requirement for dependence-based accounts, several problem cases for the transitivity thesis have emerged. Many have taken these to be counterexamples to TRANSITIVITY, but Hall [2000] has argued that we should retain TRANSITIVITY and instead take the examples to indicate that there is a problem with dependence-based accounts of causation in general.

I side with Hall in thinking that the transitivity of causation must be preserved, at least in some form. As briefly discussed in Chapter 3 §3.4 and Chapter 4 §4.4, transitivity is important to my thesis and I will argue in this chapter that such transitivity is compatible with a dependence-based account of causation, albeit not exactly as TRANSITIVITY has it. In the first part of the chapter I will show that remaining sensitive to the counterpart relations that each event is taken to fall under is the key to understanding what has gone wrong in the 'failure' of transitivity cases. The broad strategy will be familiar from Paul [2000] and Schaffer [2005]: identify illicit shifts in the middle place of the c—d—e chain. I will then consider a worry about what justification there is for reading the d event in the way that undermines the counterexamples.

In the second part of the chapter I will introduce a positive thesis: causation is only transitive when it is proportional in roughly Yablo's sense. This requires a modification either of the dependence account of causation, of the transitivity thesis, or both. I will discuss the options and offer a tentative conclusion.

7.2 Counterexamples to Transitivity

Here I consider three apparent counterexamples to TRANSITIVITY from the literature.

> **Purple Flame:**
>
> Jones puts some potassium salts into a hot fire. Because potassium compounds produce a purple flame when heated, the flame changes to a purple colour, though everything else remains the same. The purple flame ignites some flammable material nearby. Here we judge that putting the

> potassium salts in the fire caused the purple flame, which in turn caused the flammable material to ignite. But it seems implausible to judge that putting the potassium salts in the fire caused the flammable material to ignite. [Menzies, 2014][1]

In this case, in all of the closest worlds where the potassium salts are not added there remains a flame, just not a purple one. So, 'if there had been no salts then there would have been no flame' is false, but 'if there had been no salts there would have been no purple in the flame' is true. So, the potassium salts caused there to be purple in the flame, but not for there to have been a flame simpliciter.

The clean excision of the flame would avert the ignition, but simply altering the colour of the flame would not. So, 'if there had been no flame, there would have been no ignition' is true, whereas 'if there had been no purple in the flame, there would have been no ignition' is false. So, the flame caused the ignition but the purple in it did not.

Conjoining the two true causal claims you get: the salts caused the purple in the flame, and the flame simpliciter caused the ignition. There is one middle event but under two different counterpart relations: in the first counterfactual the event is essentially purple, but only accidentally a flame whereas in the second it is essentially a flame and only accidentally a purple one. So, this case does not have the format c_n caused d_p and d_p caused e_m, but rather c_n caused d_p and d_q caused e_m. Since the middle place shifts between the first claim and the second, it is not a candidate for transitivity and so cannot act as a counterexample to the transitivity thesis.

> **Dog Bite:**
>
> Terrorist, who is right-handed, must push a detonator button at noon to set off a bomb. Shortly before noon, he is bitten by a dog on his right hand. Unable to use his right hand, he pushes the detonator with his left hand at noon. The bomb duly explodes. [Hitchcock, 2001a, p.277][2]

In this case we assume that if the dog bite had not occurred, the button would still have been pressed, just not with the left hand. So, 'if there had been no dog bite then there would have been no press' is false, but 'if there had been no dog bite there would have been no left-handed press' is true. Thus the dog bite is a cause of the left-handed press but not a cause of the press simpliciter.

If there had been no press, the bomb would not have exploded, so it is true that 'the press caused the explosion'. However it is not as clear whether, if there had been no left-handed press, there would have been a right-handed press instead. Suppose that in at least some closest world there would have been a right-handed press instead and so, 'if there had been no left-handed press, then there would have been no explosion' is false. So the press simpliciter is a cause of the explosion but the left-handed press is not.

Conjoining the two true causal claims you get: the dog bite caused the left-handed press, and the press simpliciter caused the explosion. There is one middle event but

[1] This example is originally due to [Ehring, 1987, p.323].

[2] This example is attributable to McDermott [1995] but the phrasing is Hitchcock's.

under two different counterpart relations: in the first conjunct the event is essentially left-handed, but only accidentally a pressing whereas in the second it is essentially a pressing and only accidentally left-handed. So, this case does not have the format c_n caused d_p and d_p caused e_m, but rather c_n caused d_p and d_q caused e_m. Since the middle place shifts between the first claim and the second, it is not a candidate for transitivity and so cannot act as a counterexample to the transitivity thesis.

Bomb:

Billy places a bomb under a bench. Suzy goes to sit on the bench but spots the bomb and runs away instead. The bomb explodes. Suzy gets a clean bill of health the next day. So, the bomb caused Suzy's good health.[3]

In this case the clean excision of the bomb does not result in Suzy running away, but rather sitting down. So, 'had the bomb not been placed, Suzy would not have run away' is true and so the bomb is a cause of Suzy's running away.

If Suzy had not run away, she would have been blown up and so 'had Suzy not run away she would not have had a clean bill of health' is true and so Suzy's running away is a cause of her good health.

Conjoining these two true causal claims: the bomb caused Suzy to run away and Suzy's running away caused her to be in good health. This case does have the format c_n caused d_p and d_p caused e_m since there is no shift in the middle place. This case, then, is the only one of the three to formally qualify as a counterexample of the transitivity thesis.

7.2.1 Responses

Turning first to Bomb, this is the only example that has the correct structure but it will only act as a counterexample of the transitivity thesis if the result invokes TRANSITIVITY and yields an absurd result. I do not think that the result is absurd. Bombs don't cause good health, you might think, and you would be right, in general, but spotting a bomb does cause good health and you can only spot a bomb if it is there. One might be squeamish that bomb has made no difference to Suzy's health and should not count as a cause, but to hold such a line is to forget why transitivity is so important to Lewis in the first place: paradigmatic causes such as the assassin who kills the target are no less causes because the outcome was guaranteed by a back-up. Thus, they are causes even where they make little or no difference and TRANSITIVITY explains why. Bomb is a case in point.

Purple Flame and Dog Bite do give absurd results, but those results are based on a mistaken application of TRANSITIVITY. Mistaken, that is, on the reading of the middle-place event that I offered. It is essential to this outcome that the middle event shifts its counterpart relation between the first and second causal claims. In Dog Bite the essentially left-handed press must become only accidentally left-handed and in Purple Flame the essentially purple flame must become only accidentally purple.

[3] I know this example from Hall [2000], Maslen [2004a] and from Yablo [2004] but the case is widely attributed to Hartry Field (unpublished).

Discussing transitivity, Mackie [1980] points out that it is a 'very old form of fallacy' to offer 'a syllogism with an ambiguous middle term' and in recent times both Paul [2000] and Schaffer [2005] have exploited the strategy of disambiguating the middle term. Paul argues that the causal relata are event *aspects* and so the middle term *d* is a right-handed press in the first claim of Dog Bite but a press simpliciter in the second, so there is no one event aspect *d* common to both claims. Schaffer applies his contrastive account of analysing the causal claims to bring out the same difference: in Purple Flame the effect of the salts is to make the flame purple rather than not purple, but it is the fact that it is a flame rather than not a flame, that brings about the ignition. Since Schaffer takes the causal relata to be such contrastive pairs, and since there is no single contrastive pair that fits *d*, there is no case to answer for TRANSITIVITY.

However, where the contextual reading of the event is typically implicit, we can create a genuinely problematic case for my view can by making the counterpart relation for the event explicit:[4]

> **Explicit Purple Flame:**
>
> The potassium caused the purple flame (which was essentially purple and essentially a flame), and the purple flame (still essentially both purple and a flame) caused the ignition.

This case does satisfy the transitivity thesis and so the absurd conclusion that the potassium caused the ignition renders this a counterexample to either TRANSITIVITY or the counterfactual account that I am defending. (I leave it as an exercise to apply the same explicit formulation in Dog Bite). So, assuming that the counterfactual account is correct, either the transitivity thesis is false or there must be some principled reason to reject this second claim.

7.3 Proportionality

Suppose Derek's ball is scarlet and that he places it in front of Sophie, a pigeon trained to peck at all and only red things. Sophie then pecks the ball. What caused Sophie to peck the ball? Consider this causal scenario under two different descriptions:

1. The placing of the red ball caused Sophie to peck.

2. The placing of the scarlet ball caused Sophie to peck.

In the first, take redness to be an essential feature of the ball and so the clean excision of the cause event means cleanly excising the ball and the redness. Whatever replaces the ball in the closest $\neg c$-worlds will not be red and so Sophie will not peck. Thus, the clean excision standard for the cause gets the right, intuition-matching, result.

[4]Schaffer's view may have a related problem with contrived middle-place constrasts where the contrast case is 'rather than something else'. See his discussion of a boulder example in his [2005], in particular the endnotes 22-24 on p.326.

Compare this with the second description where the scarletness is flagged as essential: perhaps it is reasonable to suppose that all of the closest possible worlds in which the essentially scarlet ball is cleanly excised, no red thing takes its place. Perhaps it is, but I think it is difficult to say for sure—unlike the red case above where we can be confident—so the counterfactual test may still get the right result, but our certainty about it has shifted.

The difference between scarlet and red in this sort of case was discussed by Yablo in his [1992]. Yablo argues that the relationship of scarlet to red is that of determinate to determinable where the determinate, P, determines the determinable Q only if: (i) necessarily, for all x, if x has P then x has Q; and (ii) possibly, for some x, x has Q but lacks P [p.252]. More simply, if something is scarlet, it must be red, but if it is red it need not be scarlet. This can be translated into counterpart theoretic terms: if something is essentially scarlet, then all of its counterparts will also be red, but if it is essentially red then it may well have non-scarlet counterparts.

In the case of Sophie, Yablo points out that citing the determinate scarlet, when citing the determinable red will do, amounts to giving too much information. It need not have been that precise shade to make Sophie peck, so to be that precise about the shade is to be, if not strictly wrong, at least misleading about what was required to make Sophie peck. I may be left thinking, wrongly in this case, that my crimson ball won't illicit a peck too.

Too little information can be just as bad. Suppose that a second pigeon Trevor had been trained to peck all and only scarlet things. Does placing the red ball cause Trevor to peck? If it had not been red, then Trevor would not have pecked, so the claim looks true on a counterfactual account, but intuitively it is much better to cite the scarlet colour of the ball in explaining Trevor's peck. Being too imprecise in respect of the colour of the ball may mislead: I may be left thinking, wrongly in this case, that my crimson ball will illicit a peck too.

In the Sophie case, the scarlet was sufficient for the peck, but not required for it—it is not required because any other red would do. In the Trevor case the ball being scarlet is required, but just being red is not sufficient. Here is a proposal: for a causal claim to be properly formed the cause must be both sufficient and required for the effect. This is the essence of Yablo's proportionality constraint: the cause must be specific enough, but not too specific, with respect to the effect.

More formally, Yablo offers the following definitions:

> **Proportionality:**
>
> Where X is an event defined in terms of some property and where $+$ and $-$ indicate, respectively, more or less specificity or determinateness of the property in question.
>
> **sufficient:** X^- is sufficient for effect E iff for every X^+, if X^-, had occurred without X^+, E would still have occurred.
>
> **needed:** An event X^+, is needed for E iff for every X^-, if X^- had occurred without X^+, E would not have occurred.

This formulation adopts a fine-grained event ontology, not my preferred coarse-grained event plus counterpart ontology so I cannot adopt it as it stands, but I can translate it. In Chapter 3 (p.40) I introduced the idea of strictly more fragile and strictly more robust.[5] A reminder: Suppose that an event e can be taken to be relatively robust in context D. I will refer to its counterpart relation in D as n in that context and write e_n when referring to e under counterpart relation n. In some other context C in which e is taken to be strictly more fragile than it is when under counterpart relation n, I will refer to that counterpart relation as $< n$ (and $<< n$ and $<<< n$... for the progression of strictly more fragile counterpart relations). In some other context E in which e is taken to be strictly more robust than when it is under counterpart relation n, I will refer to that counterpart relation as $n >$ (and $n >>$ and $n >>>$... for the progression of strictly more robust counterpart relationships).

I think this allows the following analogue of the proportionality constraint, utilising event-counterpart pairs in place of Yablo's properties:[6]

> **Proportionality$_{cp}$**
>
> **sufficient$_{cp}$:** An event c under counterpart relation n—c_n—is sufficient$_{cp}$ for effect e iff for every $c_{<n}$, if c_n had occurred without $c_{<n}$, e would still have occurred.
>
> **needed$_{cp}$:** An event c under counterpart relation n—c_n— is needed$_{cp}$ for e iff for every $c_{n>}$, if $c_{n>}$, had occurred without c_n, e would not have occurred.

This way, the event of placing ball, taken as essentially red, is sufficient$_{cp}$ for Sophie's pecking since had it been crimson, and therefore not scarlet ($c_{<n}$) but still red (c_n), the ball would still have made her peck. The redness of the ball is also needed$_{cp}$ for the pecking since if the ball had been coloured $c_{n>}$, but not red c_n, the pecking would not have occurred.

The same event, taken as essentially scarlet (i.e. with counterpart relation we will call m, noted as c_m), is also sufficient$_{cp}$ for the pecking since had the ball been a lighter or darker shade of scarlet ($c_{<m}$), Sophie still would have pecked. However the essentially scarlet event (c_m) is not needed$_{cp}$ for the pecking since had the ball still been red ($c_{m>}$), but a different shade ($\neg c_m$) then the pecking would still have occurred.

On this account, the event of placing the ball is the proportional cause of the pecking when it is the placing of an essentially red ball, but not when it is the placing of an essentially scarlet ball. So far, so good.

Applied to the supposed counterexamples to TRANSITIVITY I discussed earlier, the proportionality constraint appears to support the initial reading of the counterpart relations. In Purple Flame the flame is sufficient$_{cp}$ and needed$_{cp}$ for the blaze, but the purple flame is merely sufficient$_{cp}$. Only by building too much detail into the d event description did the apparent transitivity problem appear. In Dog Bite the press

[5] Since the dimensions of fragility and robustness are inverts along the same scale, we can use *less fragile* and *more robust* interchangeably. The same applies to *less robust* and *more fragile*.

[6] I am using Weslake's 2014 paraphrase, and I alter the notation from Yablo's X^+ for more specific and X^- for less specific to my preferred reference to the robustness or fragility of the event under a given counterpart relation.

is sufficient$_{cp}$ and needed$_{cp}$ for the left-handed press, but the the left-handed press is merely sufficient$_{cp}$ for the detonation. Again, the problem lies in an overly-specific description of d.

On introducing Yablo's proportionality constraint I framed it as a constraint on a 'properly formed' causal claim. This was intentionally ambiguous between being a pragmatic or a semantic requirement. It follows, though, that if the transitivity of causation is indeed an objective, mind-independent, feature of the world, and if our causal claims need to be proportional in order that transitivity is maintained, then the proportionality constraint ought to be a semantic, not just pragmatic, constraint on our causal claims. In the latter part of this chapter I attempt to resolve that observation with my account of causation.

7.4 Failure of Transitivity in the Canonical Context

In Chapter 5 I introduced the notion of the canonical context for causal claims. This context is that idealised context in which we would give every feature of our world it's definitive (canonical) description. In such a context, I argued, the canonical description of an event would list all and only its intrinsic features and in that context no feature is more or less essential than any other. This means that in such a context events have a very strict (though not maximally strict—see p.66) counterpart relation that I refer to as i. So, event c in the canonical context assumes the counterpart relation i and is represented by c_i. I further argued that for any claim of the form 'c caused e' to be true, it must be true on the canonical counterpart relation for each event: c_i and e_i.

However, transitivity will sometimes fail in the canonical context. To see this, consider an event c with just three intrinsic features P, Q and R, an event d with just three intrinsic features S, T and U and an event e with just three intrinsic features V, W and X. Suppose further that only when P is present in c, will S be present in d and only when U is present in d, will X be present in e.

Event c is a cause of d on the canonical counterpart relation because a total excision of c_i means that no counterpart with P, Q or R will be present and so no counterpart to d_i with S will be present. Thus c_i caused d_i.

Event d is a cause of e on the canonical counterpart relation because a total excision of d_i means that no counterpart with S, T or U will be present and so no counterpart to e_i with X will be present. Thus d_i caused e_i.

Conjoining the two: c_i causes d_i and d_i causes e_i so, by TRANSITIVITY, c_i causes e_i. The problem here is that c_i causes d_i to have feature S, it does not cause it to have feature U but it is feature U, not S, that is the difference maker as to whether e_i comes to occur. In short, c had nothing to do with e. I take this to be a counterexample to the transitivity thesis. The lessons learned in the examples of Dog Bark and Purple Flame tell us where to look for the root of the problem: d is not a proportional cause of e under the canonical counterpart relation and so the essence of d is too rich and causes too much that c had nothing to do with. This generates the spurious transitive chain.

Unlike the earlier examples, proportionality cannot be achieved by refining the

counterpart relations under which the events fall. That is because the context has explicitly fixed the counterpart relation to be the canonical counterpart relation. Even for those who would deny the role I have identified for the canonical context, it remains the case that such a context is possible and so such an explicitly fixed counterpart relation is too. More generally, we can suppose that there is some non-canonical context in play, and that c, d, and e have many more than the three features I named. Nevertheless, if those features are taken to form the essence of each event, the same failure of transitivity emerges: c is a cause of d, and d of e, but c has nothing to do with e. The problem is not with the canonical context, the problem is with inflexible and out-of-proportion causes.

We could rule such causes out by insisting that a causal claim is *false* if it is not proportional. There are those in the literature who take proportionality to be a constraint on causation itself: Menzies & List [2010], Sartorio [2010] and Yablo [1992]. However, if proportionality were a constraint on genuine causal relations then that would rule out (almost) all of the canonical causal connections that I have argued are the fundamental, mind-independent relations that our common causal ascriptions track. But such a restriction would also rule out much more besides: it would be false to attribute Sophie's peck to the placing of the scarlet ball, it would be false to say 'the slamming door caused the baby to wake' or to claim that being shot by Mark David Chapman was what caused John Lennon to die. These would be false because there is some more proportional claim: that it was the placing of a *red* ball, the making of a loud noise or being shot by *anyone* that did the causal work. Perhaps such causal claims are not optimally informative of the causal structure, but that does not make them false. Imposing proportionality as a requirement on causation seems like a non-starter.

Interestingly, Weslake [2013b] argues that those who advocate a proportionality constraint on causation may be confusing causation with causal explanation. Whilst Weslake reaches this conclusion in the context of absence causation, it still fits with what I have said here. Proportionality seems to play some role in identifying the optimal form of a causal claim but that does not mean that sub-optimal causal claims are literally false. They may just be misleading or unhelpful. In the next section I suggest a less radical role for proportionality.

7.4.1 The Role of Proportionality

I take the previous section to have demonstrated that TRANSITIVITY, as currently formed, is in conflict with the counterfactual account of causation that I have been defending and so something has to give. I take it also that whilst in-proportion causal claims avoid the problem cases of TRANSITIVITY, adding a proportionality constraint to our analysis of causation itself is a non-starter. I therefore propose that we restrict the claim of the transitivity thesis to only *proportional* causes, and not causes simpliciter. This yields the proportional transitivity thesis:

> TRANSITIVITY$_p$: If c is a proportional cause of d and d is a proportional cause of e, then c is a cause of e.

TRANSITIVITY$_p$ gets each of the problem cases right, by the lights I have judged them. Purple Flame and Dog Bite do not count as cases of TRANSITIVITY$_p$ because there is no common middle-cause between the pairs of causal claims in each. Explicit Purple Flame includes an out-of-proportion second causal claim: 'the purple flame (essentially purple and essentially a flame) caused the ignition', and so does not meet the requirements of TRANSITIVITY$_p$. Even if we appealed to some pragmatic principle that could justify reading the counterpart relation of the cause to make the second claim proportional (i.e. that it is essentially a flame but only accidentally purple), it would, once again, be the case that there was no common middle-cause between the pairs of causal claims in that example. In the example which demonstrated a failure of transitivity in the canonical context, neither causal claim was proportional. Again, even if we appealed to some pragmatic principle to over-ride the explicit fixing of the counterpart relation, c would be a cause of d under some counterpart relation for d (where at least J is essential) but d would only be a cause of e under a different counterpart relation (where only L is essential).

However, TRANSITIVITY$_p$ must get both the problem cases and the straightforward cases correct to be viable. Here I consider two I introduced at the beginning:

When Billy breaks the window, his action is proportional to the breaking but Billy himself is is not needed—Suzy could have done it instead—so the proportional cause of the window breaking is that *someone* or even *something* caused the window to break. That window break, we can suppose, was not the only way that the alarm could have been triggered—a different window would have triggered it too—so the proportional cause of the alarm going off was that *some* window broke. Tweaking the middle-cause to match we get: Someone caused some window to break and some window breaking caused the alarm to go off. By TRANSITIVITY$_p$ someone caused the alarm to go off. In actuality, we know that the someone was Billy so it makes sense to say that Billy caused the alarm to go off.

In the early pre-emption case we know that the event of the rock being at that point in mid-air caused the window to break, and we know that Suzy's throwing the rock caused it to be at that point in mid-air. It needn't have been a rock of course, it could have been a ball or any other sufficiently heavy thing, so it is the event of their being a thus-and-so heavy thing in mid air that is the proportional cause of the window breaking. Similarly, it needn't have been Suzy, or indeed a person, since any projection of the heavy object would have been sufficient for it being in mid-air. So, it was the projection of the heavy object that caused it to be mid-air, and it was its being in mid-air that caused the window to break, so by TRANSITIVITY$_p$, the projection of the heavy object caused the window to break. In actuality, we know that the projection and object in question was the throwing of the rock by Suzy, so it makes sense to say that the throwing of the rock by Suzy caused the window to break.

7.4.2 ACCT and Proportion

It is not obvious how to incorporate the revised TRANSITIVITY$_p$ into the ACCT Analysis. Causal claims in the priviliged context, and subject to the ACCT Canonical causal test, will rarely be proportional since *every* intrinsic feature of the cause event is taken

to be essential in that context—many more features than will have been needed$_{cp}$ for any given effect. Yet, given that every event in the backwards lightcone of the effect will count as a cause by this standard, transitivity will rarely, if ever, be required to ensure a causal connection. Similarly the ACCT Actual causal test excises every intrinsic feature of the cause and so will rarely, if ever, be proportional in respect of the change that takes place at the effect. That suggests that only the ACCT Contextual test is really capable of generating chains of proportional causation.

It seems that any genuine case of causation between c and d will need to pass the usual ACCT Analysis tests, as will any genuine case of causation between d and e. However the conclusion that c is a cause of e is not established by these connections alone. Rather, there must be a further requirement that *some* chain of proportional causation holds between c and e. I can see no advantage in further specifying which test the proportional causal connection must hold in: if c is a cause of d and d a cause of e then in most cases we will be happy to assert causation. The counterexamples discussed here are interesting because they are exceptional. By adding in the requirement that there be at least *some* proportional causal connection between c and e we rule out the cases that gave the trouble in the first place. Refining that requirement any further would be unmotivated.

7.5 Deviant Causal Chains

The apparent transitivity of causation is often obliquely invoked when people talk of 'causal chains'. Causal theories of action and of perception have suffered counterexamples based on apparent causal chains that lead to counterintuitive conclusions. Often the problem is taken to stem from the application of causation to the theory, that is the problem is taken to relate to the theory not the causal status of the chain. The foregoing discussion of the general issues surrounding transitivity in causation motivate a reassessment of those chains in the first place. I will discuss two representative examples, one from Peacocke concerning perception and one from Davidson concerning theories of action. I aim to show that the 'deviance' of these causal chains is related to their lack of proportionality.

The first example is from Peacocke and concerns causal theories of perception which hold that to perceive a thing is to be causally related to that thing in the right sort of way. Examples such as the following put pressure on how that right sort of way might be spelled out:

> ...consider for instance the case of a man who with his eyes open but under the influence of a hallucinogen is surrounded by redwood trees that produce a scent that causes him to have a vivid visual image of redwood trees which happens precisely to match his surroundings. [Peacocke, 1979, p.123]

The thought here is that the redwood trees cause the subject to have a vivid visual image of redwood trees, just as would happen in ordinary perception, but in this case the causal chain runs via a scent and a hallucinogen. No viable theory should call this a case of veridical perception yet standard causal theories of perception do not have the

resources to say what makes the chain 'deviant' and why this doesn't count as a case of perception. Therefore, this case is a counterexample to causal theories of perception.

Most attempts to respond to this counterexample accept the causal status of the connection between the redwoods and the visual image and move on to refining or abandoning the theory which takes that connection to constitute perception. However applying the proposed TRANSITIVITY$_p$ constraint on causal chaining suggests that the causal connection is where the problem arises.

Holding fixed the presence of the hallucinogen, the scent is what gives rise to the visual image which precisely matches the surroundings. However that precise arrangements of redwoods is not a proportional cause of the scent—many other arrangements or even a synthetic scent would do the job—and so, by TRANSITIVITY$_p$, is not an appropriate link for a causal chain. That precise image was not caused by that precise arrangement of redwoods but rather was caused by there being some redwood scent present at all. Peacocke's counterexample requires that there is a perfect match between the scene and the visual image and it requires that the first causes the second. Under my proposal, this latter requirement is not met and so the counterexample fails.

Turning now to an example from Davidson which concerns causal theories of action. On such theories an event is an instance of agency if it is (at least partly) caused by a corresponding mental state within an agent. The following example is supposed to demonstrate that some further constraint is required:

> A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him as to cause him to loosen his hold, and yet it might be the case that he never chose to loosen his hold, nor did he do it intentionally. [Davidson, 1980, p.79]

Here it seems as though the climber's mental state—the desire to to be safer and the knowledge that loosening his grip would achieve this—causes him to loosen his grip, but it does so via an unintentional step. The mental state caused the action and so it can be said to demonstrate agency, on a causal theory of action. However the event was clearly lacking the sort of intent[7] required for agency and yet the causal theory of action does not have the resources to say why *this* causal chain is 'deviant'. Therefore, this case is a counterexample to such a causal theory of action.

Once again, commentators accept the causal connection between the mental state and the action and get on with refining or abandoning the theory which takes that connection to constitute action.[8] However, once again, applying the proposed TRANSITIVITY$_p$ constraint on causal chaining suggests that the causal connection is where the problem arises.

We can suppose that the nervousness caused the rope to slip but what caused the nervousness? In this instance the thoughts of letting the rope slip may be the thoughts

[7]We can simply stipulate that this is not some regular routine the climber goes through in full knowledge of the outcome. Such distinct cases do seem to demonstrate agency, as Tannjso [2009] argues.

[8]Witness the to-and-fro between Schlosser [2007] and Tännsjö [2009] in *Analysis* on this point.

that caused the nervousness, but they are only the proportional cause if there is no more robust counterpart relation of that event that would have caused the same state. What about such beliefs and wants are unnerving, we might wonder? It is surely nothing to do with ropes or weight or grip *per se*, which are the given content of the thought, but rather it is the fact that in this context that content belongs to a broader type of thought, of intentionally harming someone else, letting them fall and the consequences of such. It is that determinable type of thought, concerning harm and selfishness, and not the determinate type of thought, concerning letting a rope slip, that is unnerving. As such it is these determinable thoughts which serve as the proportional cause of the nervousness, not the overly specific, determinate, thoughts regarding ropes and slippage.

Once again the initial cause has been overly specified to create the problem. This over-specification means that the cause is not an appropriate link in a causal chain under TRANSITIVITY$_p$ so under TRANSITIVITY$_p$ the counterexample fails to demonstrate the causal link required.

These examples show that the proportionality constraint on the transitivity thesis has general application and it also shows that these cases of deviance, common across disparate areas of philosophy, admit off a general parsimonious solution. Perhaps it was a causal problem all along.

7.6 Conclusion

A counterfactual analysis requires that causation be transitive in order to correctly analyse cases of early pre-emption. In this chapter I have argued that the certain extant counterexamples to this transitivity thesis rely on a conflation: the middle-placed cause in their causal chains shifts essence mid-example. Seeking a principled reason to rule this out I have invoked Yablo's proportionality constraint for causal connectedness, but I stopped short of endorsing this as a constraint on causation in general. I have argued that the proper place for proportionality is as part of a revised transitivity thesis which holds that only proportional causation is transitive. I showed that this proposal has general application beyond the causal literature.

A worry remains however: the 'good' cases of transitivity discussed require a somewhat unnatural reading in order that they meet the proportionality constraint. It would be more elegant, and more convincing, if this reading were more naturally derived or had some deeper theoretical justification. Nonetheless, the proposal I have given resolves the problem of transitivity in a way consistent with my overall thesis.

8

Symmetric Redundant Causation

One of the key aims of my argument so far has been to demonstrate that there is a viable solution to early and late pre-emption cases—collectively cases of *asymmetric redundant causation* (ARC)—available to those who hold a counterfactual analysis of causation. Tracing the implications I have developed a view which I believe handles these asymmetric cases well, however I have said little about cases of *symmetric redundant causation* (SRC). In this chapter I will discuss the prospects of counterfactual analyses of causation in light of such cases.

In §8.1 I will first summarise the treatments of ARC cases and show that, despite their similarity, a unified solution to both is not in prospect. I will then discuss the standard cases of overdetermination discussed in the literature, offering a definition of what it takes to be a *genuine* case of overdetermination. I will then show that such cases have a distinctive counterfactual signature which can be used to set them apart as cases of SRC. In §8.2 I will show that a putative case of ARC (*trumping*), which presents a particular form of problem for any counterfactual analysis, should in fact be seen as a case of SRC and treated accordingly.

8.1 Overdetermination

8.1.1 The Anatomy of Pre-emption Cases

Recall that part of my mission in this thesis is to present an account of causation that adheres as much as possible, though not slavishly, to common sense causal judgements. Counterfactual analyses of causation do particularly well on this score when the cases are simple but by introducing unused back-ups in the vicinity of a simple causal structure such analyses struggle to track those common sense judgements. Simply put, our causal judgements clearly identify one of two candidates as a cause but

basic counterfactual analyses do not.

The 'easy' cases of pre-emption are those where one of the two putative causes triggers a causal path to the effect which cuts-off or blocks the path of the other causal candidate. By stipulating that causation is transitive one need only identify some midpoint on the causal path, after the cutting-off, upon which the effect depends. That midpoint then acts as a sort of stepping-stone of causal dependence from the effect back to just one of the causal candidates, the right one as common sense has it. I call such midpoints *partisan*. Here is an familiar example of *early pre-emption*:

> **EP:** Billy and Suzy are out to vandalise. Suzy reaches for the only rock and throws it at the window. Had she not thrown it Billy would have, and he is notoriously accurate. The rock strikes the window and the window breaks.

Intuition is crystal clear here, it is Suzy's throw that broke the window. Yet, had she not thrown, Billy would have and the window would still have broken so the breaking did not counterfactually depend on her throw. So much the worse for a simple counterfactual dependence account of causation. However, by stipulating that causation is transitive we can trace the window breaking back step-wise through some midpoint that the rock occupies in mid-air between Suzy and the window. The window breaking depended on the rock being at that point and so the rock being at that point is a cause of the window breaking. The rock being at that point further depended on it being Suzy who threw it[1] and so Suzy is a cause of the rock being at that point. Thus, by transitivity, Suzy is a cause of the window breaking via the *partisan* midpoint—the rock's position in mid-air. The right result has been secured but at the cost of marrying counterfactual theories to some sort of transitivity thesis.[2]

The 'hard' cases of pre-emption are those cases where intuition is clear on which of two putative causes is the actual cause of the effect, and where the presence of the other cause undermines dependence, but where there is no instance of 'cutting' or 'blocking' upon which to base a stepping-stone solution. Absent some un-backed-up *partisan* midpoint (upon which the effect depends, and which in turn depends uniquely on one candidate) the transitivity amendment offers no help in these *late pre-emption* (also known as *no-cutting*) cases. Common sense says one thing and theory another. Here is the example again:

> **LP:** Billy and Suzy are out to vandalise. Each throws their own rock accurately at a window but Suzy throws faster and her rock reaches the window first. The window breaks and Billy's rock sails through the void.

So, we know that Suzy broke the window but that the window breaking did not depend on her rock-throw because Billy was there to guarantee it. Had she not thrown

[1] If this isn't obvious in the example you are imagining then just add more midpoints, as many as you like, between the window and Suzy. Eventually you will have some midpoint, perhaps very close to Suzy's hand, that simply could not have come from Billy if the past is held fixed.

[2] I will not rehash the issues with transitivity here—see my discussion in Chapter 7—but suffice to say this is a commitment that complicates the theory on offer.

the rock the window would have broken regardless. It would, however, have broken later and, we can presume, differently had Billy's rock been the only one thrown. Surely that outcome is a different event from the actual breaking which happened in a specific way and at a precise time. If that is the case then the window breaking—as it actually occurred—did depend on Suzy since had she not thrown her rock *that very breaking* would not have occurred. This sort of precision in specifying the effect renders the event modally *fragile*. By altering the view of events, and not the analysis of causation itself, the counterfactual theorist can match intuition in late pre-emption cases. The foregoing chapters should make it clear how overly-simple this picture of fragility is. I have argued that fragile events are events taken to have a rich essence, where their essence, and therefore their fragility, is a context dependent matter. I have also argued that late pre-emption cases provide just the right context for the level of fragility they require. What matters here is that fragility resolves late pre-emption problems, as I argued in Chapter 3.

One might hope for a unified solution to these two cases given their common 'unused backup' structure. The fragility strategy alone, however, will not help in cases of early pre-emption since they can be constructed in such a way as to ensure the outcome effect happens at just the same time, and in just the same way, regardless of which candidate brings it about. To see this simply extend the standard early pre-emption case with an additional step whereby a periodic scan of the room triggers an alarm if the window is broken. Scanning every five minutes the system is insensitive to whether the window broke at one time (Suzy's throw) or another (Billy's throw) within that span. In either case the alarm rings at the same time in the same way. Fragility alone cannot help here.

If fragility alone cannot help in such cases, and transitivity alone cannot help in cases of late pre-emption then it seems that both tools are required to unpick the ARC cases that I have discussed. This suggests some requirements to be met by any genuine problem case for my view. It must be a case in which both causes bring about the same effect at the same time in the same way and in which neither causal path is cut or blocked-off by the other. In other words it must be a case where both causal paths 'run to completion'.[3] We do not have far to look.

8.1.2 The Anatomy of Overdetermination Cases

A putative example of such a problematic case is that of a firing squad. Assume for simplicity that they are an especially accurate and lethal squad of eight, each firing perfectly on the heart with a deadly bullet. The death of the prisoner does not depend on the first squad member, and so, by a counterfactual analysis, he is not a cause. The

[3]It is not clear how to cash out the notion of 'running to completion'. One might try to define it as being physically connected to the effect by an unbroken chain of physical connection. But this just restricts the cases the theory can handle to those worlds where events must be physically connected to their effects. Our world is, in all likelihood, such a world but our causal analysis should have wider ambitions. For a discussion of various rival definitions of 'runs to completion' and how they might impact on pre-emption cases, see Bernstein [2014]. For my part I leave the issue noted but unresolved and rely on an intuitive reading of 'run to completion' in what follows.

same goes for the second member, and the third and so on. No single member of the squad is a cause of the death and yet the prisoner surely died from being shot. This seems to mean that the counterfactual analysis has erred.

Genuine Overdetermination

First, a note of warning about such apparent cases of *overdetermination*. The death, as it actually occurred, was a death that involved the prisoner receiving damage from eight simultaneous bullets. Note that it is eight bullets, not seven or six so, on a sufficiently fragile conception of the death, it really did depend on every one of the squad: without any one of them it would have been a seven-, not eight-, bullet death. This is more than a minor quibble about the minutiae since even if we think of a more convoluted case where the death is the same no matter how many of the squad fire (perhaps they shoot at a trigger which electrocutes with even a single shot) there will still be tell-tale step on the causal path to the effect: there will be eight (not seven, or six) bullets that hit the trigger (if only one had then that would have been the cause, and the others not). There will also be significant differences in the aftermath: there will be eight bullets (not seven, not six) embedded in the trigger, and eight (not seven, not six) shells on the ground, not to mention noise, gunshot residue, remorse, payment and any other dimension along which the world diverges based on the number of shooters. I believe that this point was originally made by Bunzl [1979] who argued that there are no cases of genuine overdetermination in the actual world, just the 'illusion' of overdetermination that came from under-specifying the effect in question.[4] A sufficiently precise specification of the state of the world should show that all of the putative causes were required for the world to be precisely as it was at the time of the event, or on the path to the event, or in the aftermath.[5]

This is closely related to the fragility strategy for late pre-emption and it raises a familiar objection: the putatively overdetermined effect is not the super-fragile event that an ideal physics would describe but rather the object of a common sense causal ascription. The event picked out by 'the death of the prisoner' is not as precisely specified, or fragile, as Bunzl's account requires. In the parlance of essences, he is imputing a far richer essence (read: far fewer counterparts) for the effect than the speaker and this amounts to changing the semantic content of the causal claim made. Bunzl is probably correct to insist that, in our physics at least, the maximally fragile event did indeed depend on each of the contributions jointly. But if a more robust event (essentially a death, by shooting, at around noon) did not jointly depend on the contributions, since it would have occurred following any one of them individually, and

[4] See also Hall and Paul [2013, p.143-144]

[5] The idea that the aftermath must carry traces of the causal interaction stems from a particular aspect of our physics: that total energy must be conserved in an isolated system. If energy need not be conserved in a system then additional bullets need not create traces in the aftermath—they could simply vanish from the system entirely—however if energy must be conserved, as in our system, then additional bullets must leave additional traces in the aftermath. I will shortly consider an example which refines the problem of overdetermination by removing this contingent assumption and so removing the assumption that we can work backwards from the aftermath to determine causes. So, I will hereafter drop reference to the aftermath as it does not add anything new.

if that more robust event was what the speaker had in mind, then on what basis does Bunzl insist that the causal claim *really* meant something else? Bunzl must either give a different account of the more robust causal claim that common sense would allow, or he must hold an error theory of common sense causal ascriptions. Perhaps the latter is not so bad an option if you are happy to restrict your causal analysis to merely nomologically possible worlds, but is not in keeping with the aims of my thesis. I believe we can do better than flatly deny folk intuitions.

Bunzl's approach equates to the first of my three causal criteria, ACCT Canonical: $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$, and I think that he is correct to point out that the standard examples (such as the firing squad) are not cases of overdetermination by that standard. Unlike Bunzl I take this merely as a necessary, not sufficient, condition on the truth of a casual claim. My third criterion, ACCT Contextual, further requires that the following conditional is true: $\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n$ (where n is a counterpart relation partly determined by context). Everyday cases of overdetermination fail this criterion for any given one of the candidate causes (any given squad member). So my theory, unlike Bunzl's, reflects the intuitively problematic nature of overdetermination even in these actual world cases. My theory also lets us distinguish cases of what I will call *genuine* and (merely) *apparent* overdetermination: each candidate fails to meet the Contextual criterion in cases of apparent overdetermination but each candidate fails to meet both the Canonical and Contextual criteria in cases of genuine overdetermination.[6]

So, genuine cases of overdetermination are preciously rare if the occur at all in the actual world[7] but are they metaphysically possible? If so, they represent a problem for anyone wishing to use Bunzl's strategy in an analysis of the causal concept.

I think it is clear that they are metaphysically possible. Suppose two signals are sent down a two wires which converge. Further, imagine that the converged wire leads to a receiver which beeps if any signal is received. Both signals set off at precisely the same time, at precisely the same speed, and that each path to the signal is equally long. The receiver subsequently beeps. This looks to be a case of overdetermination but is it genuine? Perhaps not without a modification. Bunzl could point out that in our world a double-signal must arrive at the receiver, or at least at some point along the path to the signal, and that this allows us to trace the effect back to some jointly caused event (a two-signal event) which would make the activation of the receiver the product of a joint, not overdetermining, cause. However, this requires the assumption that energy is conserved as it is in our world. We need only imagine that excess signal can simply vanish, and that the converged wire can only carry the equivalent of one (possibly merged) signal, to make for a genuine case of overdetermination. In situations where the conservation of energy is not assumed, the Bunzl strategy cannot rule out cases of genuine overdetermination.

An adequate theory of causation must account for this sort of case.

[6]Actually, since failing the ACCT Canonical test entails failing the ACCT Actual test, this means that each candidate fails all three in cases of genuine overdetermination.

[7]For simplicity I am ignoring issues of compositional overdetermination (see Paul [2007]) or cases of mental-physical overdetermination. These may or may not exist regardless of whether ordinary macro-level overdetermination does and so they cross-cut this discussion.

Overdetermination Defined

In the next section I will propose a means of identifying overdetermination cases using only counterfactual apparatus. Before I do so, however, I wish to propose four joint indicators of a case of genuine overdetermination.

First of all, there must be at least two distinct events, each alone sufficient to bring about the effect. This is the basic unifying thought behind all redundant causation and I will refer to it as the *over-sufficiency* requirement.

Second, the effect must counterfactually depend upon there being at least one of the candidate events present. Without this requirement, there would be no reason to assume that the candidate events are causally related to the effect at all. Call this the *dependency* requirement.

Third, note that for it to be the *same* effect, there can be no difference in the timing or manner of the event dependent on whether one, other or both of the causes are present. If there were a difference whether one or other were present, then a fragility strategy will simply show the case to be one of *asymmetric* redundant causation instead of *symmetric*. If having both causes present made a difference over having just one, then the actual case would be one of joint causation, not overdetermination. Thus, there must be perfect symmetry in the effect no matter whether one, other or both of the causal candidates are present. Call this the *perfect symmetry* requirement.

Fourth, both causes must 'run to completion'. If one cause does not run to completion then there will be at least some point on on the path of the completed chain which is not backed-up or guaranteed by a point on the other chain. The effect would then depend upon the completed chain, and not the incomplete one and it would be a case of pre-emption and not overdetermination. Call this the *no-cutting* requirement.

If the over-suffiency, dependency, perfect symmetry and no-cutting requirements are met in a given case, then that is a case of genuine overdetermination.

Overdetermination Profiled

The problem with redundancy cases in general is that theory says a given event is not a cause whereas common sense says that it is. In asymmetric cases this is plain and obvious—Suzy causes the window to break in each construal of the story (one-rock and two-rock versions). In overdetermination cases it is not entirely clear what verdict intuition delivers: we would probably not say that either one is the cause and that the other is not, but if we say that *both* are the cause then that does not capture the notion that the effect was overdetermined rather than merely jointly-caused. Intuition is not giving theory a clear guide to the right answer here. Lewis felt this made overdetermination a special case, to be decided by the best theory (all else considered). He said:

> When common sense delivers a clear and uncontroversial answer about a not-too-far-fetched case, theory had better agree... But when common sense falls into indecision and controversy, or when it is reasonable to suspect that far-fetched cases are being judged by false analogy with commonplace ones,

> then theory may safely say what it likes. Such cases can be left as spoils to the victor, in D. M. Armstrong's phrase. [1986, 194]

Whilst there is some truth in what Lewis says here, he overlooks that whilst intuition does not give a clear positive verdict on such cases, it does at least give a negative verdict: it would be wrong to say that neither candidate caused the effect. And yet that is precisely what his counterfactual analysis (and my modification of it) say in such cases. The effect did not depend on either of the candidate causes so, taken individually, neither is a cause. This means that, if a counterfactual theory is the best theory, then we ought to deny that either of the candidates were causes. I think we can do better than this. If overdetermination cases have an identifiable structure, expressible in purely counterfactual terms, then a counterfactual analysis will be well-equipped to give overdetermination cases the status that intuition implies: it can separate them as an exception to the general theory. This is what I aim to show.

Consider the rather different case of joint-causation. Imagine that two 2kg weights are placed upon a scale which reads only up to 3kg before the screen displays an error. The displaying of the error depends on each of the weights individually and so each weight is an individual cause of the error reading. That seems wrong—it fails to capture the requirement that they both contribute—but this is not normally considered too much of a problem. The aim of an analysis of this sort is to capture what it is to be one of the great many causes that contributed and there is no suggestion that to be a 'cause' in this liberal sense is to individually suffice for the outcome. Thus, joint-cause cases are often taken to demonstrate the improvement that counterfactual theories make over sufficiency theories.[8] If we want to capture what is distinctive in joint-cause cases, all we need to do is to point out that an event is a cause of an effect if it meets the counterfactual test, and it is furthermore a joint-cause of the effect with any other synchronous event which passes the test too. In short, there is a pattern of counterfactual dependence which can be used to identify joint-causes and the existence of that pattern suggests that counterfactual dependence picks out the salient features of joint-cause phenomena.

I think a similar pattern of counterfactual dependence serves to pick out all and only overdetermining causes. Imagine again that two weights are placed on the limited scale, but this time each weighs 5kg instead of 2kg. Now the error message does not depend on either of the weights individually, but it does depend upon their conjunction. Unlike cases of joint-causation, here the individual causal candidates *fail* the initial test, but pass a subsequent conjoined test. If either event had passed the first test alone—i.e. if the effect had depended on it individually—then the other would not have been sufficient for the effect. This would violate the over-sufficiency requirement I introduced in the last section and not be a case of overdetermination at all. So, overdetermination cases must fail the first test for each individual causal candidate.

Overdetermination cases must also pass the conjoined test, lest they fail the dependency requirement (that the effect depend on their being at least one candidate cause present). However, there is a potentially problematic example for this view: perhaps

[8] This apparent improvement is discussed, and later challenged, by Hall and Paul [2013, p.148].

a 10kg weight is being held in abeyance by the presence of the two smaller weights. The smaller weights in this situation are pre-empting the 10kg weight, the presence of which guarantees the error reading. So in the absence of the smaller weights the error still appears and the smaller weights fail to jointly cause the error. However, this is just the embedding an instance of overdetermination within a pre-emption case and the earlier treatments (transitivity if there is cutting, fragility if there is not) apply. Once we consider the error event in a sufficiently fragile way we see it does in fact depend on the conjunction of the smaller weights. It seems that overdetermination cases really must pass the joint test too.

Here is the formal statement of test using a simple counterfactual analysis. For any events c, d and e, c and d overdetermine e iff: $\neg c \mathrel{\Box\!\!\rightarrow} \neg e$ is false, $\neg d \mathrel{\Box\!\!\rightarrow} \neg e$ is false, but where $(\neg c \wedge \neg d) \mathrel{\Box\!\!\rightarrow} \neg e$ is true. Such a simple counterfactual analysis is insufficient, however, in cases of pre-emption.

On my preferred counterfactual analysis there are additional steps for each causal test. However, this can be simplified here, given the perfect symmetry and over-sufficiency requirements. In genuine cases of overdetermination each cause must be individually sufficient to bring about the effect (over-sufficiency) even on the most fragile reading of the events (perfect symmetry), so each causal candidate must fail my canonical test individually. Also, by the perfect symmetry requirement, there must be no difference in the effect, even on a fragile reading, whether one, other or both of the causes are present. So, the second conjoined test (meeting the dependence requirement) for a distinctive case of overdetermination must also be run on the most fragile view of the events in question. This renders the following formal test: For any events c, d and e, c and d overdetermine e iff: $\neg Pc_i \mathrel{\Box\!\!\rightarrow} \neg Oe_i$ is false, $\neg Pd_i \mathrel{\Box\!\!\rightarrow} \neg Oe_i$ is false, and $(\neg Pc_i \wedge \neg Pd_i) \mathrel{\Box\!\!\rightarrow} \neg Oe_i$ is true.[9]

If this is correct then the ACCT theorist has the wherewithal to identify overdetermination as a special case. Where intuition hesitates, theory provides a distinctive pattern by which such cases can be isolated and set aside. Now, common sense can say what it likes about overdetermination cases and theory need only apply what it says to all and only those cases that display the requisite pattern. Overdetermination will therefore be an exception to the general causal account, but right from the start our common sense suggested they would be. I take this as a success for counterfactual theories.

[9] I note here that for any genuine case of overdetermination with two candidate causes, a third spurious cause will also pass this test. Take the weights example where the first weight is c, the second d and the error reading is e. Now, add a third event f which we can take to be the beating of some butterfly wings on the other side of the world from the scale. By the test just offered f will seem to be one of three overdetermining causes of e. That is wrong. It suggests that my formal test requires some minimal constraint in the joint test. I do not know how to offer such a test, but this is only a problem for working out which of the set that pass the test are genuine overdetermining causes, it does not affect the diagnosis of e as being overdetermined simpliciter.

8.2 Trumping

8.2.1 Pre-emption or Overdetermination?

The perfect symmetry of overdetermination cases is what makes them distinct from pre-emption cases and what makes them a distinctive type of problem for counterfactual analyses of causation. In pre-emption there must be some asymmetry in the claims of each causal candidate such that common sense agrees that one is the cause and the other is not and I have argued that when this is the case a counterfactual analyses can track common sense, given the right pragmatics and a viable account of transitivity. Any case of redundant causation in which neither candidate cause on its own would bring about the event differently, and where neither cause cuts-off the other, will meet the requirements for overdetermination that I introduced in 8.1.2. What I have said so far entails that such a case should be one of SRC however in this section I will discuss a putative counterexample to this claim from Jonathan Schaffer [2000].

Redacted Magic

In order to prime what I take to be the correct intuition in the cases, and to highlight where disagreement arises, I first offer a redacted version of Schaffer's example. I will shortly give the full version and explain the redaction.

> Suppose that at noon Merlin casts a spell (the first that day) to turn the prince into a frog, that at 6:00 pm Morgana casts a spell (the only other that day) to turn the prince into a frog, and that at midnight the prince becomes a frog. [2000, p165]

What was the cause of the prince turning into a frog? The case as it stands is under-described. We can presume that each spell is sufficient to turn the prince into a frog, but do they both do so in the same way and at the same time? If so then a difference in timing or manner of the effect may allow us to distinguish between two sufficient candidates, just as we do in cases of late pre-emption. We also do not know whether and how the spells interact. If one cuts the other off, then there may be a chain of dependence to trace as in early pre-emption cases. Schaffer fills out the details as follows:

> [T]here is neither a failure of intermediary events along the Morgana process (we may dramatize this by stipulating that spells work directly, without any intermediaries), nor any would-be difference in time or manner of the effect absent Merlin's spell[.] [2000, p165]

This means that there is an over-sufficiency for the effect, a perfectly symmetrical outcome regardless of whether one, other or both spells are cast, a dependence of the effect upon at least one of the candidate causes being present, and there is no cut-off of one spell by the other. This is a textbook case of overdetermination by my criteria in 8.1.2. Further, my counterfactual test agrees: even on a fragile reading the effect

did not depend on Merlin individually or Morgana individually but it did depend upon their conjunction.

All signs point to a case of overdetermination. At least in this redacted version.

Complete Magic

In the last section I cut off the first line of Schaffer's example and omitted the line which followed it. Here it is in full:

> *Imagine that it is a law of magic that the first spell cast on a given day match the enchantment that midnight.* Suppose that at noon Merlin casts a spell (the first that day) to turn the prince into a frog, that at 6:00 pm Morgana casts a spell (the only other that day) to turn the prince into a frog, and that at midnight the prince becomes a frog. *Clearly, Merlin's spell (the first that day) is a cause of the prince's becoming a frog and Morgana's is not, because the laws say that the first spells are the consequential ones.* [2000, p165, my emphasis to highlight the redactions.]

If we follow Schaffer's causal conclusion, this is a case of ARC, not SRC, after all. If that is right then my counterfactual test for overdetermination cases fails (it identifies a case which is not overdetermined) and the counterfactual analysis is once again faced with a case where its results and common sense diverge.

It should be apparent from comparing the redacted and completed versions of this example that Schaffer's causal conclusion is driven by the fact that it is a law of magic that the first spell cast on a given day match the enchantment that midnight. In the following section I will argue that this law is either question-begging or it is benign. In either case the law does not alter the causal structure. Trumping cases are cases of overdetermination and not cases of pre-emption.

8.2.2 Laws and Causes

In challenging the role of the magical law in the given trumping example I will give two different arguments. The first will aim to show that the stipulation of the law is question begging and the second will aim to show that a law should make no difference to our causal ascriptions if it does not make a difference to the mechanics of the case.[10]

Question Begging Laws

Suppose we had an uncontroversial case of redundant causation such as the case of late pre-emption with Billy and Suzy. Suppose further that the counterfactual dependencies are just as they are in the original set-up—had Suzy not thrown as and how she did, the window would have broken a little later, and a little differently. Suppose, however, I add to the set-up that it is a law in the world where the example takes place that the second rock thrown causes the window to break and I conclude from this that the counterfactual analysis of causation must be wrong.

[10]Similar objections to these can be found in McDermott 2002.

It is reasonable to ask: in virtue of what is it a law that the second rock thrown causes the window to break? And the answer given will need to account for the presence of 'causes' in the statement of the law. Whatever account is given, it will not be a counterfactual analysis of causation since the counterfactual dependencies render Suzy the cause on such an analysis. Whatever account is given of the introduction of the word 'causes' in the law is an account that pre-supposes the falsehood of the counterfactual analysis. As such, stipulating *that* law amounts to begging the question against the counterfactual theorist in the treatment of the example. Such laws cannot be admissible in a test case.

Schaffer does not use the word 'causes' in the formation of the law offered, but moves very quickly to taking Merlin, and not Morgana, to be the cause of the frogification "because the laws say that the first spells are the consequential ones" (ibid). If the laws did say that then the laws would be question begging, since 'consequential' is a close synonym to 'causal' in this usage. Yet the laws do not say that at all but instead say that the first spells 'match the enchantment' at midnight. This is intentionally weaker than the question-begging alternative and Schaffer goes on to argue that the law offered is not question begging.

We are asked to imagine that there are decisive competitions between the spell-casters elsewhere in the world in question such that when the spells disagree about the enchantment which is to come to pass at midnight, it is always the first spell which matches the enchantment. Assuming a supervenient account of laws along the lines of Mill-Ramsey-Lewis[11] (MRL) we take the laws of a world to be theorems of the best axiomatisation of the facts of that world. Here 'best' indicates the ideal balance between simplicity and strength, where simpler systems have fewer axioms and stronger systems convey more information about the world. Within the MRL framework we can compare two candidate laws for the situation described (assuming for simplicity that all other laws for the system are distinct from the spell-casting law):

1. The first spell cast on a given day matches the enchantment at midnight

2. When nonequivalent spells are cast the first spell matches the enchantment at midnight, but when only equivalent spells are cast, all spells match the enchantment at midnight.

According to Schaffer, the second is less simple and no more strong than the first and so, according to the MRL account, the first is a law of that world and the second is not. However, as McDermott (2002, p90) points out, Schaffer is comparing the simplicity of two *theorems* of a system, but the simplicity required by MRL is the simplicity of the *axioms*. Any world where 1 is true is also a world where 2 is true, and so any best system which has 1 as a theorem also has 2 as a theorem.

This means that we have two putative laws, each theorems of the best system, where one drives the intuition that Merlin was the cause and the other drives the intuition

[11]Such an account takes laws to supervene on facts such that facts are more fundamental than laws. This means that the laws do not dictate the facts but the other way around, which is crucial to a Humean programme that Lewis defended and that I am adopting. I understand the idea to originate from Mill [1882] and having been taken up and refined by Ramsay [1978] and [Lewis, 2001, p.72-77].

that both were a cause of the enchantment at midnight. We could even derive a third law which supported the conclusion that Morgana was the cause:

3. When nonequivalent spells are cast the first spell matches the enchantment at midnight, but when only equivalent spells are cast the last matches the enchantment at midnight.

The lesson here, I take it, is that the laws can be made to support a range of different causal conclusions. This means that selecting one candidate law above the others is to beg the question of which causal conclusion is the right one. The law Schaffer advocates begs the question in just this way and in doing so misleads intuition.

Laws and Intuition

In this section I offer two cases which I take it show that once the mechanism has been described, and once the counterfactual dependencies are made clear, the introduction of a law consistent with those counterfactual dependencies does not add relevant information to our causal judgements.

First, recall the earlier example of two signals sent down converging wires to a receiver (the version that suspended the conservation of energy assumption). This was designed to be a textbook case of genuine overdetermination. Now, suppose that the wires are coloured red and blue before they converge and purple thereafter and that on every other occasion in this world it just so happens that a stronger signal had been sent down the red wire with the result that the red-wire signal matches the purple-wire signal at the point it reaches the receiver. What does this tell us about the case when both signals have the same strength? I say: nothing. It tells us only what would have happened had the signals been different. The following may well be a theorem of the best system: 'the signal in the red wire always match the signal in the purple wire at the point it meets the receiver', but this does not make the same-signal scenario any less of a case of overdetermination.

Second, recall the case of the scale which gave an error reading for any load greater than 2kg. Adapt the scale to measure two sides against each other (as in the balancing scales of justice) and make it so that the scale reads an error if one side weighs at least 2kg more than the other (and, to avoid Bunzl-type concerns, that all excess differential pressure applied to the scale simply vanishes). Suppose there are two weights in this world, a 5kg and a 3kg one, and that they are both placed on the same side, with nothing placed on the other and that the error reading is displayed. It should be plain that this is a case of genuine overdetermination. Even though it will always be the case that whatever side the 5kg weight is on will be the side that triggers the error, it is nonetheless the case that it is no more a cause of the error reading than the 3kg weight is when they are both on the same side. This is overdetermination regardless of the law.

I think that the lesson in these cases is that the relevant counterfactual dependencies are not affected by the stipulated law. The law might inform us of competing cases, such as when Merlin and Morgana cast different spells, when the red signal is

stronger or when the 5kg and 3kg are on different sides of the scale, but it does not alter the dependence pattern of the effect on either of the candidates individually or combined. If we have an adequate treatment of cases which display such a dependence pattern—and I argued above that we do—then trumping-style cases are subject to that treatment. Similarly, if my approach to overdetermination cases is faulty, then so is my treatment of trumping-style cases. The main point here is that trumping cases are not an additional problem for counterfactual analyses of causation over and above problems of overdetermination in general.

8.2.3 Super Pre-emption

There is a further case of pre-emption that is found in the literature that I have not discussed. The problem is mentioned in passing by Hall [2004b, p.237], crediting Yablo and concerns a standard late pre-emption example with a twist. Here is the case:

> SP: Billy and Suzy are out to vandalise. Each throws their own rock accurately at a window but Suzy throws faster and her rock reaches the window first. The window breaks and Billy's rock sails through the void. Billy's rock was a *smart* rock however and it was rigged with a super-advanced guidance system. If Suzy's rock deviated from its path at any point, the guidance system would have propelled the rock to smash the window at the exact same time and in the exact same manner as Suzy's rock in fact did.

As Hall points out, the case is far fetched but he thinks that it represents the nail in the coffin of fragility-style approaches to pre-emption cases. Of course, the stipulation is such that the timing or manner of the effect event is held fixed and so a fragility strategy is supposed to be ruled out by fiat. As with other such examples, though, the devil is in the detail. I do not think this example is as problematic as Hall suggests.

Take the latest point in Suzy's rock's trajectory towards the window, the very last point on the rock's journey to break the window. Given that our world is relativistic, if this is a real-world example, or even a nomologically possible one, then if Suzy's rock is cleanly excised at that exact point then there is literally no time in which the smart-rock, however sophisticated, has to travel across the intervening space to bring about the same window breaking. If you think that there is enough time, you just haven't imagined the *very* last point on the trajectory yet. Think of a later one and try again until there is simply no time at all in which the smart-rock can travel the gap. The window breaking depended upon that latest point on Suzy's rock's trajectory and, by transitivity, on Suzy.[12]

Of course I have had to adopt an extremely fragile standard of event individuation to make this response, but Hall's case is supposed to be immune to fragility approaches and yet it isn't.

[12]Here I am assuming that there is a last point, but that just makes the exposition clearer and it isn't a strict requirement. In a continuum we do not need *the* last point, but rather some point at which the smart rock is no longer in the backwards lightcone of the window breaking, but where Suzy's is.

Maybe the case should not be restricted to the nomologically possible. Imagine a case where the smart-rock does not need any time to influence matters across the intervening space: it is capable of action at a distance. In that case we have a dependence of the effect upon not one candidate cause (Suzy's rock) but a conjunction of them (Suzy's rock and smart rock), but we have a strong preference for treating Suzy's rock as a cause of window breaking, not merely a joint-cause with some distant back-up. This is a genuine worry, but it is a worry for any situation that allows for action at a distance. Presumably there are possible worlds in which such action at a distance is commonplace, and in which such action requires no exchange of energy between the putative cause and effect. If we construct a pre-emption case in such a world, none of the strategies employed so far will be able to detect which of the various candidate causes brought about the effect.

And yet, if this is the Achilles heel of counterfactual theories, one has to wonder what the alternatives are for accommodating action-at-a-distance causation. If there is counterfactual dependence of one event upon another then a counterfactual analysis will attribute a causal relation even if those events are non-local. Without a counterfactual analysis how are we to identify cases of action-at-a-distance? Mere regularity cannot tell correlation from causation, conserved quantity theories cannot track non-local transmissions of energy and whilst contrastive approaches have their merits, they remain *counterfactual* in nature and susceptible to even the most ordinary pre-emption examples. Imperfect it may be, but the ACCT Analysis I have advocated here handles these problem cases better than all the rest.

9

ACCT, Contrastivism and Causal Modelling

9.1 Introduction

I have argued for my ACCT Analysis by starting with Lewis's original 1973 analysis of causation and proceeding to make as few amendments as possible. I began by distinguishing three challenges that any realist causal theory faces: (I) account for our everyday causal assertions; (II) give an account of the mind-independent, objective standard for causal connectedness between events; and (III) explain the relation of (I) and (II).

If the sole aim of this thesis was (I) then it would have been odd to begin with Lewis's original analysis—it has many known problems, it has many, newer, rivals and it has very few, if any, defenders left. Even within the diocese of counterfactual approaches to causation it is thought to have been superseded by the recent contrastivist accounts (of which certain modelling accounts are a sub-species). In this final chapter I aim to compare the ACCT Analysis that I have been defending with these current rivals in respect of the challenges (I) - (III) that I set out at the beginning.

9.2 Contrastivism

Contrastivism is a broader church than I will represent here. Contrastivism in general may be summarised as follows:

> A contrastivist view of a concept holds that all or some claims using that concept are best understood with an extra logical place for the contrast class. [Sinnott-Armstrong, 2012, p.134]

So, contrastivism about causation holds that all or some claims using the causal concept are best understood with an extra logical place for the contrast class. In the case of causation there remains a question of whether a single contrast class is required for the cause or the effect, or whether two contrast classes are required: one for each. I think Schaffer [2005, 2012a] has argued convincingly that contrastivists about causation ought to apply a contrast class to both the cause and the effect in causal claims. Therefore, I will only consider quaternary contrastivist accounts of causation as relevant here. According to such an account 'c is a cause of e' is true iff c rather than C^* causes e rather than E^*, where what counts as a member of C^* or E^* is a function of context.[1]

Such accounts can be found in Maslen [2004a], Schaffer [2005, 2012a], Northcott [2007] and List & Menzies [2010] and whilst I will apply the most charitable rendering of contrastivism I can to each of the problems discussed below, the fullest case, and the richest account of the benefits of contrastive causation, comes from Schaffer's [2005]. I will therefore default to Schaffer's contrastivism unless I state otherwise.

9.2.1 Everyday Causal Talk

I will begin by considering the advantages that contrastivists claim to offer over the original Lewis theory in respect of our everyday causal talk and then consider how the ACCT Analysis compares.

Contextualism: Our causal attributions vary with context so that the same causal sentence in different contexts can have different truth conditions. The original Lewis analysis has no context-variant element in the semantics which would explain this variation but contrastivism has: context shifts the relevant contrast class for each of the events in question. Given the right contrastivist reading, the causal claim thereby shifts truth conditions across contexts.[2]

Absences: We are perfectly happy to say that an absence of rain caused the crop to fail or that the absence of precaution caused the fire and whilst a standard Lewisian counterfactual account of causation gives corresponding truth conditions for these assertions, it requires an ontology of omissions—events that are defined negatively as the absence of something else. This is metaphysically abhorrent [Schaffer, 2005, p.330]. The Lewisian might say that the absence of precaution is just identical with the watching of T.V. that occurred instead, but there are two problems with this idea: it takes the watching of T.V. to have caused the fire, which seems wrong, and it locates the cause at a spatiotemporal distance from the effect without there being any oomph that travels the intervening space. The contrastivist can instead take the negative nominal depicting the event—i.e. 'the absence of precaution'—to pick out the watching of T.V. in the actual world but as also picking out the taking of precautions in some other world. This avoids the implication that watching the T.V. rather than reading

[1] Some notation: following Northcott [2007] I will denote actual events as c and e and associated contrast classes (non-empty sets of contrast events, possibly singleton sets) using C^* and E^* respectively. Note that this differs from Schaffer's specific contrast cases, c^* and e^*. Here I am aiming for a more wide-reaching discussion and so I use the more widely used version involving contrast sets.

[2] I omit discussion of the 'Selection' here for brevity. See Chapter 2 §2.4.1.

a newspaper caused the fire. In that other world the precautions are spatiotemporally local to the region that caught fire in the actual world.

Transitivity: The original Lewis theory does not seem to have the resources to diagnose what goes wrong in cases where transitivity appears to fail (see Chapter 7). If throwing potassium salts into a fire causes it to burn purple, and if the purple flame causes the curtains to catch fire, then it seems as though throwing potassium salts into the fire caused the curtains to catch fire. The conclusion seems false and a contrastivist can say why: the middle place of the transitive chain conflates two distinct claims: in the first conjunct the salts causes there to be a purple flame rather than an orange flame, but in the second conjunct it is the fact that there is a purple flame rather than no flame that causes the curtains to catch. The rather-than clauses expose the illicit shift that has taken place. Lacking those clauses, Lewis's original account does not have the resources to reject these counterexamples.

Regarding Contextualism, my ACCT Analysis offers a rival account whereby counterpart variation, i.e. shifts in the modality we attribute to the events involved, accounts for the shifts in truth value. In Chapter 2 I argue that this sort of contextualist view is at least as good as contrastivism at handling the examples and that it does so without multiplying argument places beyond necessity.

Absences are metaphysically abhorrent, "bogus entities" as Lewis [2004a, p.100] put it. Endorsing absence causation as genuine raises problems of profligacy, non-locality and location and my discussion of contrastivist responses in Chapter 6 should make it clear that extant contrastivist approaches may improve on the original Lewis account, but they do not improve enough. My ACCT Analysis, which takes absence causal claims to be literally false, instead glosses our acceptable causal assertions as assertions about would-be, not actual, causal relations. As such, they serve a useful explanatory role without being literally true claims about our world. I think this account is an improvement on the contrastivist story.

The contrastivist does have a neat story to tell about transitivity, however. The illicit shifts in the middle place do seem like the source of the problem in the standard putative counterexamples. My ACCT Analysis alights on the same point but I rely on a proportionality constraint to rule out certain problem cases (see Explicit Purple Flame, Chapter 7). The problem with this reliance is that it pushes me to choose between considering proportionality as a constraint on the literal truth of causal claims or as a constraint on the assertability or well-formed status of certain causal claims. Whilst the first option rescues the plausible idea that causation itself is transitive, it comes at the cost of ruling out a load of our ordinary causal claims which are not strictly proportional (such as that Mark David Chapman firing a gun caused John Lennon to die). The second option gives up on transitivity of causation itself, but saves the idea that some causation—namely the proportional sort—is transitive. This is not as neat as the contrastivist picture as it stands but this is due to the ACCT Analysis commitment to there being an objective causal relation. This discussion is supposed to leave that issue aside for the purposes of comparing the relative success of contrastivism and ACCT approaches vis-á-vis ordinary causal talk. The ACCT Analysis introduces a proportionality constraint as a constraint on causal talk involving chains and the

contrastivist introduces a matching-contrast constraint to apply to causal talk involving chains. The contrastivist and ACCT theorist alike can eschew such constraints for causal statements concerning direct causal influence, so it seems that both make a special case of talk involving indirect chains.

One area where there appears to be no progress offered by the contrastivist is in dealing with the problems of late pre-emption and overdetermination. Schaffer explicitly accepts this [2005, p.358, note 35] but whilst few others even mention it, no-one seems to think that contrastivity offers an improvement over Lewis's 73 analysis in dealing with this central problem. Of course I have argued at length that adopting a fragile view of events and a sensible pragmatics of causal discourse, helps resolve this issue. Leaving aside the maximally fragile events that I endorse for the purposes of identifying an objective causal relation (more on this shortly), the contrastivist could simply help themselves to the pragmatic maxims that I have offered modulo a little terminology. I offered three maxims:

1. Take the effect to be as fragile as is required to render the claim true.

2. Take the effect to be no more fragile than is required to render the claim true.

3. Take the effect to be only as fragile as the speaker could discern or infer at the time.[3]

Fragility may not have a direct correlate in certain contrastivist theories. Schaffer, for example, explicitly rejects fragile events on the basis that he can simply specify the target event in the actual world and a sufficiently close alteration in the contrast class where required. Others, such as List & Menzies would perhaps be able to retain the fragile terminology. In any case, the contrastivist has the resources to render the desired 'Suzy's throwing the rock caused the window to break' as true even when Billy throws too (albeit slightly later): Suzy's throwing the rock rather than dropping it, caused the window to break as it did, rather than break slightly later. My argument in Chapter 3 was supposed to justify this sort of reading on the basis that it was charitable and proportional to the information available in the situation. Although no contrastivist has yet argued this line, I see no reason why they could not avail themselves of such a pragmatic approach.

So, concerning the first aim—to account for our every day causal assertions—I take the honours to be roughly even. Whilst I think it is a mistake for contrastivists to endorse absence causation as genuine causation, and whilst no contrastivist has yet argued that they can resolve late pre-emption problems, these are merely contingent failings and there remains logical space for a contrastivist theory which does not have these shortcomings. Perhaps such a contrastivist position would have the slight edge given their neater treatment of counterexamples to transitivity.

[3]Recall that this last maxim trumps the first in cases where they conflict. See Chapter 3, §3.2.3.

9.2.2 Objectivity

My second requirement for a theory of causation was that it give an account of the mind-independent, objective causal relation between distinct events in the world. Here the ACCT Analysis gives a clear answer: c is a cause of e simpliciter iff $\neg Pc_i \ \Box\!\!\rightarrow \neg Oe_i$. Causal contrastivism takes causal claims to be four-place relations where two of those places (the contrast classes) are typically determined by the context. Contextual theories thus seem at first glance to be ill-suited to giving an account of such an objective relation, especially a causal relation that seems to relate two events, rather than four.

However, causal contrastivists, facing the problem of objectivity, may wish to endorse something like my two-tiered system of analysing the genuine causal relation in the world on the one hand and analysing the content of our causal talk on the other. If they so wished, I think they should say something like the following: Causal assertions take place in ordinary contexts and so the contrast places (C^* and E^*) are populated by the salient alternatives which the context implies. Claims about the mind-independent, objective matters of fact regarding causation should be taken to occur in a privileged context in which the contrast places (C^* and E^*) are populated by the maximum number of alternatives to the actual c and e—i.e. every possible alteration of c and e.

Some details matter here. First, the maximal set of alternatives must be the maximal set of alternatives *that differ intrinsically* from the actual c and e. Otherwise, there will be contrast cases that are identical to the actual world in all respects local to the causal interaction under consideration—there would be no difference and so there could be no difference makers to be discovered by the contrastive test. In effect, to be an alteration of c or e requires being at least not intrinsically identical to them.

Second, and more importantly, somewhere in the totality of other worlds there will be worlds where some member of C^* occurs and some member of E^* does not (or vice versa). This will be true for any C^* and E^* simply because every contingent possibility is manifested in the total set of possible worlds. If the contrastivist was committed to every member of the contrast set of C^* being paired-off with some member of the set of E^* then every causal claim would be false because there are distant worlds where some C^* manifests but some E^* does not (or vice versa).[4] So instead such a contrastivist should commit only to there being a pairing of the closest C^* and E^* alternatives in the privileged context.

With these two stipulations in place, there remains a telling difference between my ACCT Analysis and that of the contrastivist and this difference stems from the asymmetric standard of non-occurrence that I argued for in Chapter 4. In the privileged context, and with the above stipulations in place, the contrastivist could offer the following objective standard of causation:

> c is a cause of e simpliciter iff: c occurs in the actual world, e occurs in the actual world and, in all of the closest possible worlds in which no intrinsic

[4] Schaffer seems to entertain something like this idea in his [2005, p.348] but is careful to distance himself from a commitment to it.

> duplicate of c occurs (C^*-worlds), no intrinsic duplicate of e occurs (E^*-worlds).

Whereas I would instead endorse:

> c is a cause of e simpliciter iff: c occurs in the actual world, e occurs in the actual world and, in all of the closest possible worlds where c is cleanly excised ($\neg Pc_i$-worlds), no intrinsic duplicate of e occurs ($\neg Oe_i$-worlds).

Notice that on this notion of contrastivism, the set of the closest C^*-worlds include worlds where some very close alternative to c occurs, maybe an alternative that alters some aspect of c (say, the shade) in a way that makes no difference to e. If such an alteration of c is possible, then the contrastivist in the privileged context would have to consider that c is not a cause of e simpliciter, even for paradigmatic causal cases such as that of a spark causing a fire. By my ACCT Analysis, on the other hand, the closest $\neg Pc_i$-worlds do not include such close variants of c. This means that my view will not suffer from such false negatives.

Can such asymmetry be built into the contrastivist system? I think it can. It requires that only clean excisions are considered genuine alterations of the cause event. The same will not apply to the effect contrast set as I argued in Chapter 4 and so the contrastivist system can be rendered asymmetric by applying a different standard to the contrast sets on either side of the causal equation. With this final alteration in place the contrastivist can match the ACCT results in the privileged context and endorse an extentionally equivalent test for genuine causal relatedness between distinct events in the world.

Note, though that if the contrastivist endorses this asymmetry, as I say they must, they give up their neat account of transitivity. The same cases with the same problems will apply when the objective causal test is applied: suppose that the clean excision of c alters one aspect of d but only the alteration of a different aspect of d can affect e. So, c is a cause of d, and d is a cause of e, but nothing about c impacts anything about e. This is a structural, not intuition-based, counterexample to the transitivity of causation itself and it applies as much to the contrastivist as it does to me. Whatever advantage of neatness the contrastivist could claim in relation to my (I) objective is now wiped out.

It is worth noting that adopting a fine-grained notion of events may reduce the number of instances of this type of counterexample, though I do not think it would rule them out altogether. The problems stem from events having multiple 'aspects' such that the middle event in a transitive chain can have one aspect caused by c and go on to cause e in virtue of another, separate, aspect. Thus, adopting a fine-grained notion of events would appear to avoid such problems so long as it counted different events for each aspect.

I think this would cure some cases, but not all. A fine-grained view of events could of course count more events, and discriminate more finely, than a course-grained view in the privileged context. However if there are any complex events—events constituted by more than one property or aspect—then such an event can be used to create the

same type of transitivity issues that I introduced. Given that complex events will be vital if a theory is to accommodate our ordinary causal notions of 'Suzy's throwing a rock' or 'Derek's placing the red ball', I find it difficult to imagine a theory of events that avoids this structure altogether.[5]

What is more, the positive cases for adopting a course-grained view of events (with counterparts) is strong. Taking events to be regions with context-variable counterpart relations has helped account for the contextual variance of our causal claims. It has also given us a language in which to highlight and discuss the implicit shifts in modality of the causes and effects in pre-emption cases and thereby enabled us to express the entailment principles behind the pragmatic maxims that help to clarify these centrally problematic cases. Perhaps this work can all be translated into a fine-grained events ontology but I do not know how to do so and I would not know how to demonstrate the crucial relation between our everyday talk and the objective standards for causal connectedness in such an ontology.

Which leads me neatly onto my (III) aim for the thesis: show how our everyday causal talk relates to the proposed objective standard for causal connectedness.

9.2.3 Connection

Suppose that the contrastivist has taken on my suggested privileged context approach to the objective standard of causal connectedness. How does such a pliable contrastivist's account of everyday causal claims and their account of the objective standard relate to one another? Is the everyday talk left floating free of the objective truth?

I do not see that there is any reliable connection between the two. Any causal claim that is true in an ordinary context will utilise the same actual-world target events (c and e) as are under consideration when the same 'c caused e' claim is considered within the privileged context. What alters with the context is the contrasts. The ordinary claim will imply a comparatively specific alternative (Suzy dropping the rock, for example) whereas the same 'c caused e' claim being considered in the privileged context implies a far wider contrast class. Given the closeness requirement (see above), only the closest such contrasts will be relevant to the objective status of c and e. I think this creates a problem which can be teased out in the following example. Suppose that Billy throws his rock after the window is broken. Billy is not a cause of the window breaking since in all the closest alterations where Billy's throw is cleanly excised, the window nevertheless breaks. However, if the contrast case is a situation where Billy had thrown a minute earlier (ahead of Suzy) then the following contrastive claim would come out true: Billy's throwing the rock when he did *rather than* a minute earlier caused the window to break when it did *rather than* break earlier.

This shows that it is possible for there to be a true contrastive claim in an ordinary context which does not match the truth in the privileged context. This is possible because in the ordinary context there is no requirement that the alternative under

[5]It is worth noting that the classic transitivity cases in the literature are introduced against a background expectation that the counterfactual theorists for whom they cause a problem are, like Lewis, fine-grainers. Fine-graining seems unlikely to help in these cases.

consideration in C^* be one of the closest alternatives. Thus, the contrastive standards come apart.

In summary, whilst the contrastive approach does well in handling our causal talk, and whilst an objective causal standard can likely be expressed using the contrastive resources, under a contrastive theory there does not seem to be the same close connection between the truth of what people say and an objective standard of causal connectedness between events. By comparison, the ACCT Analysis meets all of the desiderata.

9.3 Causal Modelling

Rivalling this sort of four-place contrastivism are a range of interventionist causal modelling approaches. Whilst I have not discussed these theories so far in the thesis, they are influential and current and I think that they deserve consideration. A full critique of causal modelling is a topic for another thesis, but here I will aim to sketch some of the similarities and differences between a sophisticated causal modelling approach and my preferred ACCT Analysis.

Versions of causal modelling can be found in Sprites, Glymour and Scheines [2000], Hitchcock [2001a] and Halpern and Pearl [2005] (following Pearle [2000]) but I believe the most comprehensive and philosophically sophisticated account was given by Woodward's seminal book *Making Things Happen* [2003]. Given its status and the great interest that it has generated, I will focus almost exclusively on Woodward's account but the broader issues will be common (in one form or another) across all extant views of causal modelling. I will begin with some preliminary explanation of the modelling project and then compare Woodward's account with the ACCT Analysis on offer just as I did with contrastivism.

9.3.1 Causal Modelling Introduced

Modelling accounts have in common the idea of forming structural equations between variables and defining causal connections between the variables in terms of those equations. The pairing of the Variables ($\mathcal{V}$) and the Equations ($\mathcal{E}$) forms the model $\mathcal{M}$, so $\mathcal{M}$=[$\mathcal{V}$,$\mathcal{E}$]. Whether or not one variable is a cause of another is determined by whether a certain kind of intervention performed upon the first variable alters the state of the second variable. Thus, it is a form of counterfactual dependence theory, where the dependencies are encoded in the structural equations in particular and therefore in the model in general. Theories differ on what the appropriate variables might be, what makes for an appropriate model, which variables should be included and what exactly counts as an intervention. There is also disagreement about what the ambition of such a programme is—to define causation or causal explanation.

Setting those issues aside for the moment, consider this simple example. Suppose that Suzy is alone and throws a rock at a window. The rock hits the window and the window smashes. In this scenario we have three variables ST (Suzy throws), SH (Suzy's rock hits the window) and S (the window smashes). Variables can take a

(possibly continuous) range of values but in this case let us stipulate that Suzy can throw hard (ST=2), Normal (ST=1) or not at all (ST=0). That the rock either hits the window (SH=1) or it does not (SH=0) and that the window either breaks (S=1) or it doesn't (S=0). Let me also state that no matter how hard it is hit, the window will break if struck at all and that in the actual situation Suzy threw the rock hard (ST=2), it struck the window (SH=1) and the window broke (S=1).

Now we can form the structural equation for this scenario. Those variables which are given a specified value are named *exogenous* and those whose value derives from the others in the model is *endogenous*. For each endogenous variable in the model there is an equation that takes that variable on the left hand side and some equation for working out its state on the right. For the simple example given, here are the equations for the case described:

ST = 2 (exogenous)

SH = 0 if (ST=0), 1 otherwise (endogenous)

S = SH (endogenous)

This model therefore specifies every possible state that every variable will take once ST is known. This means that we can determine what value S would take if an intervention was performed to alter ST from ST=2 to ST=1. Under such an intervention the resultant value would remain S=1 so no matter whether Suzy throws hard or normally the window still breaks. *If* a causal modelling theory required that for ST to be a cause of S *every* alteration of ST would have brought about *some* alteration of S, then since this intervention on ST made no difference to S, Suzy's throw would not be a cause of the window breaking. This is clearly false.

Why this is false is perhaps familiar from my discussion in Chapter 4. We do not care that there is some feature of the cause that has no impact on the effect, we care that there is some other feature of the cause which does. Causal modelling theories recognise this and typically take a variable X to be a cause of variable Y just in case there is some alteration of X in the model that changes the value of Y. So, since intervening on ST such that ST=0 alters S from S=1 to S=0, ST is a cause of S, even if no other alteration of ST made any difference to S. Suzy is a cause of the window breaking after all.

Two other concepts will be required for the forthcoming discussion. Woodward's theory (which I take to be representative and to which I will default to on detail), employs the notion of a *direct cause* (DC) and the notion of a *path*. Intuitively a direct cause is an unmediated cause but here is Woodward's more precise definition:

> DC: A necessary and sufficient condition for X to be a direct cause of Y with respect to some variable set $\mathcal{V}$ is that there be a possible intervention on X that will change Y (or the probability distribution for Y) when all other variables in $\mathcal{V}$ besides X and Y are held fixed at some value by interventions. [2003, p.55]

The reason for the clause about fixing all the other variables is that you may intervene on ST in our model (ST=0) and that will change the value of S (S=0) but ST is not an unmediated cause of S because the rock must first hit the window or not as represented by variable SH). On Woodward's careful definition ST is not a direct cause since holding SH at its value in the actual case (SH=1), no intervention on ST makes a difference. (It is worth noting for later the oddity of Suzy's rock hitting the window but her not throwing it—that is the sort of situation under consideration and the oddity of some permissible combinations of variables will be relevant when discussing pre-emption below). Interventions on SH on the other hand do vary S even when every other variable is held fixed, so SH is a direct cause of S (and ST is a direct cause of SH too).[6]

Intuitively a directed path is the route or chain by which a cause impacts on the effect. More precisely a path is a series of direct causes that runs from X to Y. The notion of a path is used to define what it is to be a contributing cause: X is a contributing cause of Y so long as there is some manipulation of X that alters Y when *all off-path variables in the model are held fixed at their actual values*.[7]

Much more detail is required to do justice to the depth and complexity of causal modelling approaches in general and Woodward's in particular. Here, though, I will stick to the points that I need and bring out detail where it matters. Before I begin to look at how this account fares on the issues of contextualism, transitivity, absences and redundancy, I first want to point out some features of this approach that relate to the questions that were raised in this thesis.

The first issue to note is that the causal relata need not be events. Whatever can take a value of a variable in the model is ripe to be considered a cause. The variables in the model might be 'Suzy's throw' or 'the velocity of the rock' or 'the mass of the rock' or there might be a single variable which captures all of this information at each given value: if Suzy throws a 4kg rock at 15m/s the value of the ST variable is ST=3, if she throws a 2kg rock at 15m/s the value of the ST variable is ST=27, and so on. Alternatively the ST variable could have just two values: throws (ST=1) and does not throw (ST=0) ignoring the different ways each state might manifest.

The point here is that causal modelling is strictly compatible with a further requirement that only events can occupy the variable slots, though it would seem to be unmotivated from the point of view of tracking counterfactual dependencies. Equally, it could be the case that the only causal claims we are interested in concern coarse-grained events which would be represented as sets of variables set at a value. Thus the event c is not represented simply the value of one variable (X) but rather by a range of values of variables $[\mathcal{V}_1 = v_1, \mathcal{V}_2 = v_2, \mathcal{V}_3 = v_3 \ldots \mathcal{V}_i = v_i]$. So, causal modelling is compatible with the assumption that coarse-grained events are the causal relata, but causal modellers do not restrict themselves to that notion.

[6]Weslake [forthcoming, p.6] seems to think that as long as X appears on the RHS on the equation for Y, X is a direct cause of Y. This might prove neater but the case for such a variation on the DC definition is not made.

[7]This appears to be weakness in Woodward's programme. Weslake [forthcoming] argues that by holding all the off-path variables fixed at their actual values, simple pre-emption cases get the wrong result. A full and detailed discussion of this issue is found in his [forthcoming].

Notice, though, that there is a choice to be made when setting up the model as to how to represent the occurrences in the world. As I showed with ST above, it can take a wide range of values which capture a lot of information about the world under a single variable (ST) or the model could simply capture a binary value (1,0) for that variable (ST). Each element in the world could, in theory, have its own variable or it could be categorised into a smaller list of variables (Throw, Velocity, Mass), each of which take one of a larger range of values. The point here is that the model could have fewer variables, each with many values, or it could have more variables with fewer values and the choice here is not dictated by the way the world is but rather by how we choose to represent it in the model. Woodward is clear on this issue: "... conclusions about causal relationships are sensitive to one's choice of representation." [2003, p.80] So, different contexts may well require different assignments of values and variables.

Bearing this in mind, reflect on the definition of a direct cause. A direct cause is one *unmediated* by other variables in a given model. And a contributing cause is further defined in terms of paths of direct causes. Thus whether something is a contributing cause or a direct cause is a model-relative matter. I will say more on this in the next section but it is a striking feature of causal modelling approaches in general.

Finally, consider the notion of an intervention in Woodward's theory. An intervention is an idealised manipulation of the actual values where at least one variable is set to a non-actual value regardless of whether that variable was endogenous or exogenous in the set-up of the model. This makes a great deal of sense when we look at the directed graphs that are typical in modelling texts. These graphs link vertices (representing variables) and use arrows to depict direct causal connections between the variables. An intervention that alters the value of a variable effectively 'severs' the inward arrows on the graph to indicate that the intervention ensures that variable is no longer a function of any other. This importantly relates to Lewis's 'miracles'. For Lewis, when we consider what would have occurred under some modification or other of the actual occurrences we should consider those closest worlds in which the modification has been made. Here closest is playing a substantial role as it requires us to rank worlds with alterations by their similarity to the actual world. Lewis [1979] gives some guidelines about how we should do this but Woodward [2003, p.134–137] criticises Lewis's criteria as under-motivated and ambiguous. According to Woodward we should think of interventions as a special kind of idealised causal contribution that alters the value of some variable and holds fixed certain others. One might well agree with Woodward's complaints about Lewis's system but it is important to note that by characterising an intervention in causal terms, as Woodward does, the resultant theory of causation is explicitly non-reductionist. Direct causes and contributing causes are defined in part by the notion of an intervention, and an intervention is in turn defined in causal terms. We have a circular definition of causation. Woodward is adamant that the circularity is not vicious and that it remains illuminating but we need not disagree with him on either of these points to conclude that something desirable has been compromised by this approach: namely a reductive account of causation.

9.3.2 Everyday Causal Talk

Having set the prelimenaries, I will now compare the prospects of a Woodward-style causal modelling approach to the ACCT Analysis I have advocated in light of my desiderata (I)–(III). First, to the requirement that a viable theory ought to account for our ordinary causal claims. Again, I will take each variety of problem discussed in the thesis and consider how causal modelling theories and the ACCT Analysis compare in handling these cases.

Contextualism: Causal modelling theories define variables as causes relative to a model where the model in question is influenced by the contextually salient factors. If context indicates that the salient difference is between serving awkwardly and serving gracefully (see Chapter 2), stealing rather than not stealing or drinking hemlock rather than not, then the model will reflect that by including these variables and the appropriate values of those variables. The truth of a causal claim can thereby shift with context, as the examples discussed in Chapter 2 would suggest.[8]

Absences: Modelling theories are concerned only with variables and values of variables and so no ontological commitment to the nature of absences is required. If the plants being watered is a variable that can take values 1 or 0, then that variable set to 0 appears to indicate an absence or an omission. If shifting the value of that variable from 0 to 1 alters the value of some other variable (plants dying) in the model (holding off-path variables fixed) then the first variable is a cause of the second. So, causal modelling can endorse absence causation as long as the model is appropriately set up without having any trouble with the issue of *locating* the absence somewhere in the world.

Causal modelling approaches can also reject the notion that the Queen's failure to water the plants caused them to die, since the context will not licence including the Queen (or Obama or a velociraptor) as a variable in the model. This avoids the problem of *proliferation*.

The problem of *non-locality* remains, however. If preventing Bomber at point A entails some alteration to the subsequent state at point B, then the causal modeller will endorse full-blooded causation between the two occurrences (in an appropriate model) and so the Bomber has an impact at a spatiotemporal distance and disconnect. This remains problematic.

Transitivity: Woodward's approach does not require that causation be transitive and is not committed to the counter-intuitive consequences of the counter-example cases [p.57-58]. Nevertheless, some causes influence their effects in a mediated way such that if the cause (X) had been varied, but some mediating variable (Z) along the path to the effect (Y) had been held fixed, then X would not be considered a cause of Y. Thus Woodward instead endorses the notion of a contributing cause whereby only the off-path variables are held fixed [p. 59]. For an extended discussion of this issue see Weslake [forthcoming]. On the whole, causal modelling approaches seem to have the resources to give the right results in the problem cases so long as the model is

[8] I do not think that causal modelling approaches need be committed to a *contextualist* semantics to account for this variation—a *relativist* semantics seems viable too—but it will streamline the discussion to assume a contextualist reading.

appropriately set up.

Redundant causation: cases of pre-emption and overdetermination are problematic for Woodward's initial formulation of actual causation in just the same way as they are problematic for Lewis: the effect is insensitive to each cause individually. This motivates a further modification of Woodward's account of what the off-path variables should be set to when we are considering the causal status of mediated variables (i.e. variables that influence the effect via another). A reminder of the late pre-emption case:

> **LP:** Billy and Suzy are out to vandalise. Each throws their own rock accurately at a window but Suzy throws faster and her rock reaches the window first. The window breaks and Billy's rock sails through the void.

The causal modelling treatment of this case is to model Suzy's throw (ST), Billy's (BT), Suzy's rock hitting the window (SH) and Billy's (BH), and the state of the window (W). What this approach is able to demonstrate is that if you hold fixed the fact that the backup (Billy's rock) did not strike the window (BH=0), then the smashing of the window (W) is sensitive to interventions on Suzy's throwing of the rock (ST) and hence Suzy is a cause of the window's breaking. Exactly how you specify which off-path variables are held fixed remains unclear in these cases, however. What is it about Billy's rock striking the window that picks that variable out to be held at its actual value?[9]

Further, Hall [2006] argues that there is no state of the model that actually gets the cases of late pre-emption correct. The apparent success of causal modelling in these cases hinges on the idea that BH is held fixed at value 0. Leaving aside the important issue of when to fix off-path variables, and at what values, BH=0 is ambiguous across two different states of the world: one in which Billy's rock vanishes, and another where it passes through the space the window occupied (a moment after the window smashed). This seems innocuous enough at first, after all causal modelling approaches do not distinguish many ways in which a variable might take a value. However, Hall argues that there is no *stable* disambiguation of BH=0 which yields the desired result in the late pre-emption example. Suppose that BH=0 represents the rock not being in the vicinity of the window, either because Billy didn't throw it or because it has vanished courtesy of an intervention. Suppose further that BH=1 represents Billy's rock striking the window and BH=2 represents Billy's rock being in the same place as BH=1 but where the window is not present. This successfully disambiguates the state of Billy's rock. Now, in the actual pre-emption case BH=2 by this schema since Billy's rock is in the right place to smash the window just Suzy got there first. Now, if Suzy smashes the window first, then W=1, but if she doesn't smash the window first, and if BH is held fixed at BH=2, then how can the window not be smashed and yet Billy's rock pass through the space it occupies? Plainly it cannot. From this line of reasoning,

[9] I again direct the interested reader to Weslake [forthcoming] who discusses a range of options and problem cases for Woodward's, Hitchcock's, and Halpern and Pearl's existing attempts to resolve this problem.

Hall concludes that late pre-emption cases have not yet been resolved by the adoption of a casual modelling approach to causation.

Even if we suppose that Hall is right about the existing approach, I think the causal modeller can instead apply a model which contains enough fine discrimination between variable so as to ensure that asymmetric cases of redundancy (pre-emption) no longer give the wrong value. What I have in mind here is just the fragility strategy by a different name: if X and Z are candidate causes for the window breaking and if X pre-empts Z, then there will be a model such that *window breaks at time* t_1 is one variable and *window breaks at time* t_2 is another. X is a cause of the first such variable and deserves the title of cause, just like Suzy in the classic example.

Genuine overdetermination remains untouched by any fine-graining of the model (otherwise Bunzl-style objections re-emerge as discussed in Chapter 7 §8.1.2).

I think the contextualism of the model and of my counterpart relations are in fact related. The counterpart relation that applies to an event in a context tells us under what conditions that event will be taken to occur and which it will not. This is functionally equivalent to fixing the values that can be assigned to the variable that represents that event in any causal model—it fixes the way the world must be for the value 1 to apply to the 'event occurred' variable. I think this suggests that the causal modeller and ACCT theorist are tracking the same contextual features.

Regarding absences, however, the causal modeller accords positive and negative causes equal weight. In fact the causal modeller accords actual causes and would-be causes equal weight in general, which I take to be problematic. A person claims something rather different when they say that X did cause the accident than they do when they say that X could have. In the first case we consider what did happen and entertain counterfactuals about it (the ACCT Analysis and causal modelling theories are alike in this regard). However, in the second case we consider what would have occurred if I had acted in such and such a way and consider counterfactuals about that would-be situation. I think this is an important distinction, and not just at the level of normativity where responsibility is apportioned, but also at the level of causation simpliciter. I had nothing to do with the accident in the second case but a causal model which takes the variable *Neil Intervenes* (and values 0 for no, 1 for an ordinary sort of intervention and 2 for an extremely unlikely intervention) will still consider me a cause of the accident, even if the only way I could have intervened was by some very unlikely and difficult process.[10] The problem here is that would-be actions are considered on a par with manifest actions in a causal model but not in our common sense reasoning about causation. I take this is a weakness of causal modelling views.

The ACCT Analysis and Woodward's version of interventionism share the denial of the (straightforward) transitivity of causation.[11] However the ACCT Analysis, unlike Woodward's, seeks to explain why we thought that it was. Here I think the score is not yet settled, however, as there are rival accounts of how to define which off-path

[10]The modeller can, as always, insist that their model would exclude the more improbable option that makes the claim so implausible, but I think this is just to highlight how dependent our causal attributions are upon a seemingly fickle process of model creation.

[11]Recall that in Chapter 7 I denied that causation itself was transitive but argued that *proportional* causation was.

variables are to be held fixed when considering mediated causes. I remain optimistic on behalf of the causal modeller, however, as they have the expressive resources to say anything that the ACCT Analysis can about when chaining occurs and when it doesn't, and they can always add a transitivity clause to the theory if that is what turns out to be best.

Pre-emption cases generate similar issues for the causal modeller regarding chaining, but as I pointed out above, the appropriate model (i.e. one with the right variables and the right values) can yield the right result. Some account of how to read the context in the right way to justify such a model would need to be forthcoming, however. I offer the beginnings of such a pragmatics in Chapter 3 but as far as I can tell, no comparable account exists for the causal modeller.

Overdetermination cases cannot be resolved by interpolating variables, however. In cases of genuine overdetermination, the end variable takes a specific value no matter whether one, other or both of the putative causes and set to their actual values and no amount of fine-graining the effect variable or value will tell the alternative combinations apart. Notice though, that the distinctive pattern of counterfactual dependence that I appealed to in Chapter 8 can be expressed just as easily in a causal modelling framework. Where neither counts as a cause on its own, but where their conjunction does, two variables can be defined as overdetermining the effect. This does not fall on one side or the other of considering such overdetermining variables causes or not but, as I argued previously, this is perhaps an advantage given that intuition is unclear as to what the right answer is in these cases. I think the following from Woodward can be read as supporting this approach:

> My guess is that Lewis is wrong about common sense [regarding overdetermination], but it also seems to me that in an important respect it does not matter much whether we count c1 and c2 as causes of e in this case as long as we can agree about what the relevant patterns of counterfactual dependence are. [Woodward, 2003, p.85]

As with the comparison with the contrastivist, I take the honours to be roughly even in this comparison. Held to the standard of everyday talk the ACCT Analysis and causal modelling approaches can track contextual shifts in the semantic value of a causal claim, account for absence causal attributions, avoid the counterexamples regarding transitivity and offer a way of understanding redundant causation. I think it is a straightforward weakness that the causal modelling approaches cannot distinguish actual from would-be causation and, by extension positive from negative causal claims, but I think that the expressive resources of a causal modelling theory which tracks context (variables, values, paths and interventions) are extremely powerful and that we should be hopeful on future progress on the outstanding issues when the target concept is contextual.

9.3.3 Objectivity

The second requirement for my ACCT Analysis was that it give an account of the mind-independent objective relation between events that was the causal relation. Here

the ACCT Analysis gives a clear answer: c is a cause of e simpliciter iff $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$. Causal modelling, like contrastivism, is inherently context-embedded and also seems at first glance to be ill-suited to giving an account of such an objective relation.

Woodward's version has no pretensions of reduction but, once a model is established, the truth conditions for causal claims are a matter of objective fact. To this extent the theory does give us an objective-ish standard of causal relatedness between variables. For the purposes of explanation and the functioning of science, Woodward argues, this is objective enough [p.56-57].

Perhaps this sense of objective is enough for a causal explanation, but it is not enough for the committed causal realist who seeks an account of the mind-independent relation of causation. Considered as a theory of causal explanation, I have little to complain about on this score—to my mind it is just as context and interest-dependent as explanations ought to be—but as a standard of causation I think this fails the objectivity requirement in its current form.

It remains open to the causal modeller, just as it did to the contrastivist, to endorse a two-tier system of analysing causation along the lines of the ACCT Analysis: one account for our talk and its explanatory role, and a second account of the genuine relation of causation in the world that is independent of interest. The account given by Woodward clearly fits the mould for the first sort of analysis but in order to meet the requirements of the second sort of analysis there would need to be some singular model that captured our world. Schaffer seems to agree:

> But given that different models yield different causal verdicts, and given that there is no unique notion of a canonical model for a given situation (at least none yet developed), it might seem that the only remaining option is to relativise causal relations to models. [Schaffer, 2014]

Could there be a canonical model? I think my argument from Chapter 5 in favour of positing a privileged context suggests that there could. Models are relative to contexts and so for there to be a canonical model there would need to be an appropriate context in which every causal fact was captured. Such a context would be idealised one, and one that stands out from all other contexts—it is a privileged context. In the privileged context, I argued, every intrinsic feature of every region is relevant and so would need to be represented in the model. A model which captures every detail and dependency in the world would indeed deserve the title 'canonical'.

There are two questions to ask of the canonical model: does it capture all of the causal facts? Does it capture only the causal facts? I will focus on the second question here without intending to prejudice the first.

By Lewisian standards, we assess a counterfactual conditional $\neg c \boxdot\!\!\rightarrow \neg e$ by checking whether all of the closest possible worlds in which the antecedent is true are also worlds where the consequent is true. Without the restriction to only the closest of all possible worlds every contingent counterfactual conditional would be false since there will always be some world where the antecedent is true and the consequent false. No matter how odd that world might be, it remains within the scope of all worlds and so to get meaningful truth conditions for our counterfactual conditionals, we need to

restrict the worlds we consider. How to do this is indeed problematic, but Lewis gives us an account to start with.

In trying to give a truly objective standard of causation, Woodward has a parallel problem. If there is some non-actual value of X which would generate a non-actual value of Y, then according to Woodward's theory X causes Y (details aside for the moment). Given that this theory allows X to be a cause of Y if any state of X would alter Y, then if the model contains every possible value for X and not just some restricted interest-relative set of values, then there will be some extremely improbable or remote possible values that X could take. Suppose X represents the mass of some earthly object that does not affect Y at all (it just outside the backwards light cone, say). An idealised intervention on X to set the mass to that of Jupiter will indeed alter the value of Y. This issue can be re-created indefinitely for a great many obvious non-causes. When the model is relative to an ordinary context, such outlandish possibilities are prohibited by the intuitive irrelevance of the possibility. However in the canonical model we require a mind-independent standard by which we can restrict the values to only the realistic or relevant ones. Lewis's similarity criterion plays the role in his schema and that approach has been roundly criticised by Woodward, but no equivalent is offered for the causal modelling theory.

This is not exactly a criticism of Woodward's theory of course since the avowed aim of that theory is to give an account of causal explanation and causation-in-context. Nevertheless, I think that it is a highly significant benefit of the ACCT Analysis over causal modelling approaches that it can give a plausible account of the mind-independent, objective relation of causation in our world.

9.3.4 Connection

Suppose for the sake of argument that there were some canonical model and that the theory incorporating it somehow ruled out the far fetched and problematic cases. How would that account of the mind-independent, objective causal relation connect with the account of our ordinary talk?

I think the first thing to notice is that the canonical model would obviously endorse many more causal relations in the world than a restricted, context-relevant, model. Take a model which represents the firing of a gun (FG) and the death of a senator (DS) and assume that if FG=1 then DS=1 too, but DS=0 otherwise. In the canonical model there will be many more variables such as the pulling of the trigger (PT), the motion of the hammer (H), the expulsion of gasses (EG) and so on that constitute the act of firing the gun. We can also interpolate as many additional variables between FG and DS as there are points in space between them. If FS is a cause of DS via a directed path of such variables, then every variable on the path is a contributory cause, too. Adding more variables adds more causation so the canonical model will posit more causal relationships than a restricted and contextualised model will.

Perhaps the connections that count as causal on the contextualised model are a strict subset of those that are causal on the canonical model. If so we should expect that when two variables are causally connected in a contextualised model, those same variables will be causally connected in the canonical model. The following from

Woodward would seem to support that thought:

> [A]s long as there is a single causal rouite from X to Y, if X is a contributing cause of Y, X will remain a cause of Y (although not a direct cause of Y) if additional variables Z are interpolated between X and Y along this route. Thus, in the example above, if A's pulling the trigger is a contributing cause of B's death with respect to a variable set that does not include the release of the spring, the hammer striking the cartridge, and so on, it remains a contributing cause when the variable set is expanded to include these variables. [Woodward, 2003, p.56]

The reverse is obviously false: you cannot infer from X being a cause of Y in a model with many variables that X will still be considered a cause of Y on a less detailed model. The model may not even include X or Y as variables! So this truth preserving interpolation is asymmetric and it favours adding more detail and finer discriminations rather than the opposite. I believe this idea is closely related to the asymmetry of precision I discussed in Chapter 3. In that case I argued that tending to a fragile interpretation meant that you tended to truth. Woodward seems to be endorsing something rather similar.

A problem awaits, however. If the causal facts of a contextualised model are to be a strict subset of the causal facts of the canonical model, then there had better not be any causal connection that is taken to exist in the contextualised model but not in the canonical model. However there is nothing to stop the variables in the contextualised model being variables that do not appear in the canonical model at all. Perhaps the variables representing firing a gun and the death of the senator are macro-level variables that simply do not appear in the set of micro-level variables that constitute the canonical model. If so there is a causal connection in one model that does not appear in the other. Of course, if these macro-level variables reduce to micro-level variables then there will still be a close connection between the two, but there is no guarantee that every model will reduce in the appropriate way. There may be variables that represent wildly disjunctive or gerrymandered properties in which only one of those properties is considered a cause in the canonical model. How could we reduce such a variable to its components without being taken to posit false causal connections?

Once again this is a problem which has a parallel in the Lewisian approach. Lewis has to stipulate that excessively extrinsic or disjunctive events cannot stand in a causal relation. In doing so he is imposing a restriction on what the causal relata could be. The causal modeller traditionally avoids specifying what the variables need to be in their theory. Typically seen as a strength, perhaps this secular approach hinders the causal modeller's ability to draw a tight logical connection between the canonical and contextualised models.

I have taken the Woodward theory quite far off the reservation in the hope that it could be rendered metaphysically satisfactory by the lights I set in my introduction. I cannot say that the trip was successful but it seems that there is scope to express some of the same notions, and be open to some of the same problems and objections,

as the ACCT Analysis. Perhaps some future version of a causal modelling theory will be able to meet all of my desiderata, but for the time being the best candidate remains the ACCT Analysis I have introduced and defended in this thesis.

10
Conclusion

In this thesis I have been arguing on the assumption that a viable theory of causation should account not only for our causal talk, but also for the causal relation in the world. Sceptics who deny that there is such a relation the world [Menzies, 2009, Menzies & Price, 1993] will of course reject some of the reasoning that I have deployed in my case for the ACCT Analysis of causation, but the viability of such a realist theory stands as reason to think that those sceptics have given up on the naturalist project too soon. They may continue to question the motivation for the theory I have offered, but the viability of that theory undermines the motivation to be sceptical in the first place.

I too gave up swiftly on the idea that there was a single account to be given that would match both desiderata but I did not give up on the hope that our causal talk and the causal facts of the matter were related in some intimate way. The ACCT Analysis that I have offered assigns the same basic counterfactual structure to the conditions for true causal talk and for determining the brute causal facts concerning distinct events c and e. Event c is a cause of e relative to counterpart relation x iff c and e are linked by a chain of causal dependence, where e causally depends on c iff $\neg Pc_x \ \Box\!\!\rightarrow \neg Oe_x$ is true. In this rendition x is a free variable representing the counterpart relation under which events c and e fall. In ordinary contexts where our everyday causal talk takes place, the variable x is set to a context-sensitive value n representing the counterpart relation that applies in that context. This relativises the truth of the conditional, and so the truth of ordinary causal assertions, to contexts and interests in a way that is at odds with a realist causal project. I argued in Chapter 5 that the genuine causal relation should be taken to be determined in a privileged context and that in such a context a canonical counterpart relation applies. This is signified by replacing x with i in the counterfactual conditional above. Thus, a two-tier approach was proposed: one tier for ordinary causal talk (where x is replaced by n) and another tier for the genuine causal relation where (x is replaced by i).

So, we were left with two tests with a common structure, but how do they relate?

I first tried to show that our causal assertions about c and e could only be true if c and e were causally connected in the privileged context, that is: for all n $(\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n)$ only if $(\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i)$. However, this attempt failed for an instructive reason: in cases of prevention or would-be causation, the first can be true and the second false. I deem prevention and would-be causation to be non-actual causation—a controversial but precedented stance—and so this prompted me to introduce a further condition on our causal talk to establish the actual causal connection between c and e in a way that $\neg Pc_n \boxdot\!\!\rightarrow \neg Oe_n$ did not. This second condition on our causal talk required that the following conditional (or chains thereof) be true: $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_n$. Given that this final condition can only be true if $\neg Pc_i \boxdot\!\!\rightarrow \neg Oe_i$ is true (see Chapter 9, §5.4), this extra condition ensures that, by my theory, our causal talk asymmetrically depends upon the causal facts. This is as it should be.

There are a cluster of objections that are commonly levelled against simple counterfactual analyses of causation and I have discussed what I take to be the most problematic of these: contextual variation, late pre-emption, absence causation, failures of transitivity, trumping pre-emption and overdetermination. In addressing contextual variation in Chapter 2, I only required the resources of a counterpart theory of events to rescue a simple Lewisian theory of causation (I called the combination CCT) and this same counterpart theory allowed me in Chapter 3 to show a problem in Lewis's reasoning around late pre-emption cases and his influential rejection of fragile events. Using basic pragmatic maxims, I argued that a flexible standard of fragility was justified by the context and could resolve the late pre-emption issues. This pragmatic approach highlighted a hitherto unnoticed asymmetry in our standards of non-occurrence for cause and effect and in Chapter 4 I argued that this asymmetry be built into our causal analysis. Having argued that we needed a two-tier account using this asymmetric standard (which I call ACCT) in Chapter 5, I then addressed the issues of absence causation, transitivity and trumping/overdetermination with these new resources. In Chapter 6 I argued that absence causation was not genuine and should be considered would-be causation, i.e. causation centred on another world, in Chapter 7 that causation was not transitive but that proportional causation likely was, and in Chapter 8 that trumping cases were just overdetermination cases in disguise. Overdetermination itself acquires a distinctive counterfactual profile on my ACCT Analysis and so whatever verdict intuition confers on such cases can be matched by fiat if required. Such cases, and those of the more problematic action-at-a-distance cases are very-far fetched given our physics, however.

I have not established that the ACCT Analysis is true, of course. That was never the aim. The aim was to 'measure the price' as Lewis put it and I considered the relative price of contextualist and causal modelling alternatives in Chapter 9. In that discussion I showed that the ACCT Analysis could match many of the benefits of the other theories at the level of our causal talk but that it was in a unique position with respect to my realist assumptions. Of the views under consideration only my ACCT Analysis could meet the desiderata that we (I) account for our everyday causal assertions; (II) give an account of the mind-independent, objective standard for causal connectedness between events; and (III) explain the relation of (I) and (II).

Lewis alluded to such a two-tier project in his Postscripts to Causation in [1986e,

p.199] when he said ‘To say how the double standard works may not be a hopeless project, but for the present it is not so much unfinished as unbegun.’ Having taken great inspiration from his own work on the philosophy of causation, and having found my own path within his neo-Humean framework, I hope I have shown that it is a far from hopeless project and that it is, at last, begun.

Bibliography

Achinstein, P. (1975). Causation, transparency, and emphasis. *Canadian Journal of Philosophy*, *5*(1), 1–23.

Anderson, P. (1972). More is different. *Science*, *177*(4047), 393–396.

Anscombe, G. E. M. (1957). *Intention*. 40. Harvard University Press.

Barker, S. (1999). Counterfactuals, probabilistic counterfactuals and causation. *Mind*, *108*(431), 427–469.

Batterman, R. (2013). The tyranny of scales. In R. Batterman (Ed.) *Oxford handbook of the philosophy of physics*, (pp. 256–286). Oxford University Press.

Beebee, H. (2004). Causing and nothingness. In Collins et al. [2004], (pp. 291–308).

Beebee, H. (2009). Causation and observation. In Beebee et al. [2009], chap. 22, (pp. 471–497).

Beebee, H., & MacBride, F. (2014). De re modality, essentialism, and lewis's humeanism. In B. Loewer, & J. Schaffer (Eds.) *Blackwell Companion to David Lewis*. Blackwell Publishers: New York.

Beebee, H., Menzies, P., & Hitchcock, C. (Eds.) (2009). *The Oxford Handbook of Causation*. Oxford University Press.

Bennett, J. (1988). *Events and Their Names*. Hackett.

Bernstein, S. (2013). Omissions as possibilities. *Philosophical Studies*, *167*(1), 1–23.

Bernstein, S. (2014). A closer look at trumping. *Acta Analytica*, (pp. 1–22).

Björnsson, G. (2007). How effects depend on their causes, why causal transitivity fails, and why we care about causation. *Philosophical Studies*, *133*(3), 349–390.

Broadbent, A. (2012). Causes of causes. *Philosophical Studies*, *158*(3), 457–476.

Bunzl, M. (1979). Causal overdetermination. *Journal of Philosophy*, *76*(3), 134–150.

Coady, D. (2004). Preempting preemption. In Collins et al. [2004], (pp. 325–340).

Coates, P. (2000). Deviant causal chains and hallucinations: A problem for the anti-causalist. *Philosophical Quarterly*, *50*(200), 320–331.

Collins, J. (2000). Preemptive prevention. *Journal of Philosophy*, *97*(4), 223–234.

Collins, J. (2009). Counterfactuals, causation, and preemption. Http://collinsjohn.org/ccp.pdf.

Collins, J., Hall, N., & Paul, L. (Eds.) (2004). *Causation and Counterfactuals*. MIT Press.

Davidson, D. (1963). Actions, reasons, and causes. *Journal of Philosophy*, *60*(23), 685–700.

Davidson, D. (1967). Causal relations. *Journal of Philosophy*, *64*(21), 691–703.

Davidson, D. (1969). The individuation of events. In *Essays in Honor of Carl G. Hempel*, (pp. 216–34). Reidel.

Davidson, D. (1970). Events and particulars. *Noûs*, *4*(1), 25–32.

Davidson, D. (1980). Freedom to act. In *Essays on Action and Events*, (pp. 63–82). Oxford University Press.

Dowe, P. (2000). Causality and explanation. *British Journal for the Philosophy of Science*, *51*(1), 165–174.

Dowe, P. (2001). A counterfactual theory of prevention and causation by omission. *Australasian Journal of Philosophy*, *79*(2), 216–226.

Dowe, P. (2004a). Causation and misconnections. *Philosophy of Science*, *71*(5), 926–931.

Dowe, P. (2004b). Causes are physically connected to their effects: Why preventers and omissions are not causes. In C. Hitchcock (Ed.) *Contemporary Debates in Philosophy of Science*, (pp. 189–196). Blackwell Pub.

Dowe, P. (2004c). Chance-lowering causes. In P. Dowe, & P. Noordhof (Eds.) *Cause and Chance: Causation in an Indeterministic World*. Routledge.

Dowe, P. (2008). Causal processes. In *Stanford Encyclopedia of Philosophy*. Stanford: The Metaphysics Research Lab.

Dowe, P. (2009a). Absences, possible causation, and the problem of non-locality. *The Monist*, *92*(1), 23–40.

Dowe, P. (2009b). Causal process theories. In Beebee et al. [2009].

Dowe, P. (2009c). The power of possible causation. Http://philsci-archive.pitt.edu/4768/.

Dowe, P. (2009d). Would-cause semantics. *Philosophy of Science*, *76*(5), 701–711.

Dowe, P. (2010). Proportionality and omissions. *Analysis*, *70*(3), 446–451.

Eells, E. (2002). Propensity trajectories, preemption, and the identity of events. *Synthese*, *132*(1-2), 119–141.

Ehring, D. (1987). Causal relata. *Synthese*, *73*(2), 319–328.

Ehring, D. (2009). Causal relata. In Beebee et al. [2009], chap. 19, (pp. 387–413).

Eklund, M. (2001). Supervaluationism, vagueifiers, and semantic overdetermination. *Dialectica*, *55*(4), 363–378.

Fazekas, P., & Kampis, G. (2012). Turning negative causation back to positive. Http://philpapers.org/archive/FAZTNC.pdf.

Field, H. (2003). Causation in a physical world. In Beebee et al. [2009], (pp. 435–460).

Fine, K. (1994). Essence and modality. *Philosophical Perspectives*, *8*, 1–16.

Fraassen, B. C. V. (1980). *The Scientific Image*. Oxford University Press.

Frisch, M. (2010). Causes, counterfactuals, and non-locality. *Australasian Journal of Philosophy*, *88*(4), 655–672.

Funkhouser, E. (2009). Frankfurt cases and overdetermination. *Canadian Journal of Philosophy*, *39*(3), 341–369.

Garrett, D. (2009). Hume. In Beebee et al. [2009], (pp. 73–91).

Glynn, L. (2013). Of miracles and interventions. *Erkenntnis*, *78*(1), 43–64.

Grice, H. P. (1989). *Studies in the Way of Words*. Harvard University Press.

Grice, H. P. (2013). 4. logic and conversation. In M. Ezcurdia, & R. J. Stainton (Eds.) *The Semantics-Pragmatics Boundary in Philosophy*, (p. 47). Broadview Press.

Grice, P. (1968). Logic and conversation. In *Studies In The Way of Words*. Cambridge: Cambridge University Press.

Hájek, A. (2002). Counterfactual reasoning (philosophical aspects)—quantitative. In N. J. S. . P. B. Baltes (Ed.) *International Encyclopedia of the Social and Behavioral Sciences*, (pp. 2872–2874). Elsevier. Manuscript.

Hall, N. (2000). Causation and the price of transitivity. *Journal of Philosophy*, *97*(4), 198–222.

Hall, N. (2002). Non-locality on the cheap? a new problem for counterfactual analyses of causation. *Noûs*, *36*(2), 276–294.

Hall, N. (2004a). The intrinsic character of causation. *Oxford Studies in Metaphysics*, *1*, 255–300.

Hall, N. (2004b). Two concepts of causation. In Collins et al. [2004], (pp. 181–276).

Hall, N. (2006). Comments on woodward, "making things happen". *History and Philosophy of the Life Sciences*, *28*(4), 611–624.

Hall, N. (2007a). Structural equations and causation. Available from: http://nrs.harvard.edu/urn-3:HUL.InstRepos:3710361.

Hall, N. (2007b). Structural equations and causation. *Philosophical Studies*, *132*(1), 109–136.

Hall, N., & Paul, L. A. (2003). Causation and preemption. In P. Clark, & K. Hawley (Eds.) *Philosophy of Science Today*. Oxford University Press.

Hall, N., & Paul, L. A. (2013). Metaphysically reductive causation. *Erkenntnis*, *78*(1), 9–41.

Halpern, J. Y., & Pearl, J. (2005). Causes and explanations: A structural-model approach. part i: Causes. *British Journal for the Philosophy of Science*, *56*(4), 843–887.

Hart, H., & Honore, A. M. (1959). *Causation in the Law*. Oxford: Clarendon.

Hawthorne, J. (2005). Chance and counterfactuals. *Philosophy and Phenomenological Research*, *70*(2), 396–405.

Hitchcock, C. (2001a). The intransitivity of causation revealed in equations and graphs. *The Journal of Philosophy*, *98*(6), 273–299.

Hitchcock, C. (2001b). A tale of two effects. *Philosophical Review*, *110*(3), 361–396.

Hitchcock, C. (2004). Do all and only causes raise the probabilities of effects? In Collins et al. [2004], (pp. 403–418).

Hitchcock, C. (2007). Prevention, preemption, and the principle of sufficient reason. *Philosophical Review*, *116*(4), 495–532.

Hitchcock, C. (2009). Causal modelling. In Beebee et al. [2009].

Hitchcock, C. (2013). What is the 'cause' in causal decision theory? *Erkenntnis*, *78*(1), 129–146.

Hitchcock, C., & Knobe, J. (2009). Cause and norm. *Journal of Philosophy*, *106*(11), 587–612.

Hitchcock, C. R. (1995). Salmon on explanatory relevance. *Philosophy of Science*, *62*(2), 304–320.

Hitchcock, C. R. (1996). The role of contrast in causal and explanatory claims. *Synthese*, *107*(3), 395–419.

Honoré, A. (2008). Causation in the law. In E. N. Zalta (Ed.) *Stanford Encyclopedia of Philosophy*, 46, (p. 92). Stanford: The Metaphysics Research Lab.

Hume, D. (1975). *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*. Clarendon Press, 3rd ed.

Hume, D. (1978). *A Treatise Of Human Nature*. Oxford: Clarendon Press, 2nd ed.

Jago, M., & Barker, S. (2012). Being positive about negative facts. *Philosophy and Phenomenological Research*, *85*(1), 117–138.

John Collins, L. P., Ned Hall (2004). Counterfactuals and causation: History, problems, and prospects. In Collins et al. [2004], (pp. 1–59).

Kim, J. (1973). Causation, nomic subsumption, and the concept of event. *Journal of Philosophy*, *70*(8), 217–236.

Kim, J. (1974). Noncausal connections. *Noûs*, *8*(1), 41–52.

Kim, J. (1976). Events as property exemplifications. In M. Brand, & D. Walton (Eds.) *Action Theory*, (pp. 310–326). D. Reidel.

Knobe, J. (2009). Folk judgments of causation. *Studies in History and Philosophy of Science Part A*, *40*(2), 238–242.

Kratzer, A. (1977). What 'must' and 'can' must and can mean. *Linguistics and Philosophy*, *1*(3), 337–355.

Kratzer, A. (1981). Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, *10*(2), 201–216.

Kratzer, A. (1989). An investigation of the lumps of thought. *Linguistics and Philosophy*, *12*(5), 607–653.

Kratzer, A. (2005). Constraining premise sets for counterfactuals. *Journal of Semantics*, *22*(2), 153–158.

Kratzer, A. (2009). On the plurality of verbs. Http://semanticsarchive.net/Archive/jI4YWRlO/PluralityKratzer.pdf.

Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press.

Kvart, I. (2001a). The counterfactual analysis of cause. *Synthese*, *127*(3), 389–427.

Kvart, I. (2001b). Lewis's 'causation as influence'. *Australasian Journal of Philosophy*, *79*(3), 409–421.

Kvart, I. (2004). Causation: Counterfactual and probabilistic analyses. In Collins et al. [2004], (pp. 359–386).

Latham, N. (1987). Singular causal statements and strict deterministic laws. *Pacific Philosophical Quarterly*, *68*, 29–43.

Lavelle, J. S., Botterill, G., & Lock, S. (2013). Contrastive explanation and the many absences problem. *Synthese*, *190*(16), 3495–3510.

Lewis, D. (1968a). Counterpart theory and quantified modal logic. *Journal of Philosophy*, *65*(5), 113–126.

Lewis, D. (1968b). Counterpart theory and quantified modal logic. *Journal of Philosophy*, *65*, 113–26.

Lewis, D. (1973). Causation. *Journal of Philosophy*, *70*(17), 556–567.

Lewis, D. (1979). Counterfactual dependence and time's arrow. *Noûs*, *13*(4), 455–476.

Lewis, D. (1983a). Counterparts of persons an their bodies. In *Philosophical Papers*, vol. I, (pp. 47–54). Oxford University Press.

Lewis, D. (1983b). Holes. In *Philosophical Papers*, vol. I, (pp. 3–9). Oxford University Press.

Lewis, D. (1983c). *Philosophical Papers*, vol. I. Oxford University Press.

Lewis, D. (1983d). Postscripts to counterpart theory and quantified modal logic. In *Philosophical Papers*, vol. I, (pp. 39–46). Oxford University Press.

Lewis, D. (1983e). Scorekeeping in a language game. In *Philosophical Papers*, vol. I, (pp. 233–249). Oxford University Press.

Lewis, D. (1986a). Causal explanation. In *Philosophical Papers*, vol. II, (pp. 214–240). Oxford University Press.

Lewis, D. (1986b). Events. In *Philosophical Papers*, vol. II, (pp. 241–269). Oxford University Press.

Lewis, D. (1986c). *On the Plurality of Worlds*. Blackwell Publishers.

Lewis, D. (1986d). *Philosophical Papers*, vol. II. Oxford: Oxford University Press.

Lewis, D. (1986e). Postscripts to causation. In *Philosophical Papers*, vol. II, (pp. 172–213). Oxford University Press.

Lewis, D. (2001). *Counterfactuals*. Blackwell Publishers.

Lewis, D. (2003). Things qua truthmakers. In H. Lillehammer, & G. Rodriguez-Pereyra (Eds.) *Real Metaphysics: Essays in honor of D. H. Mellor*, (pp. 25–38). Routledge.

Lewis, D. (2004a). Causation as influence. In Collins et al. [2004], (pp. 75–106).

Lewis, D. (2004b). Void and object. In Collins et al. [2004], (pp. 277–290).

Lipton, P. (1991). Contrastive explanation and causal triangulation. *Philosophy of Science*, *58*(4), 687–697.

Lipton, P. (2007). Alien abduction and inference to and best explanation. *Episteme*, *7*, 239.

Livengood, J., & Machery, E. (2007). The folk probably don't think what you think they think: Experiments on causation by absence. *Midwest Studies in Philosophy*, *31*(1), 107–127.

Mackie, J. L. (1980). The transitivity of counterfactuals and causation. *Analysis*, *40*(1), 53–54.

Maslen, C. (2004a). Causes, contrasts, and the nontransitivity of causation. In Collins et al. [2004], (pp. 341–357).

Maslen, C. (2004b). Degrees of influence and the problem of pre-emption. *Australasian Journal of Philosophy*, *82*(4), 577–594.

McDermott, M. (1995). Redundant causation. *British Journal for the Philosophy of Science*, *46*(4), 523–544.

McDermott, M. (2002). Causation: Influence versus sufficiency. *Journal of Philosophy*, *99*(2), 84–101.

McGrath, S. (2005). Causation by omission: A dilemma. *Philosophical Studies*, *123*(1-2), 125–48.

Mellor, D. H. (1995). *The Facts of Causation*. 1. Routledge.

Mellor, D. H. (2004). For facts as causes and effects. In Collins et al. [2004], (pp. 309–23).

Menzies, P. (1988). Against causal reductionism. *Mind*, *97*(388), 551–574.

Menzies, P. (1989). Probabilistic causation and causal processes: A critique of lewis. *Philosophy of Science*, *56*(4), 642–663.

Menzies, P. (1996). Probabilistic causation and the pre-emption problem. *Mind*, *105*(417), 85–117.

Menzies, P. (1999). Intrinsic versus extrinsic conceptions of causation. In H. Sankey (Ed.) *Laws and Causation: Australasian Studies in the History and Philosophy of Science*, (pp. 313–329). Kluwer.

Menzies, P. (2002). Is causation a genuine relation? In H. Lillehammer, & G. Rodriguez-Pereyra (Eds.) *Real Metaphysics: Festschrift for D. H. Mellor*. Routledge.

Menzies, P. (2004). Difference-making in context. In Collins et al. [2004], (pp. 139–180).

Menzies, P. (2007). Causation in context. In H. Price, & R. Corry (Eds.) *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Oxford University Press.

Menzies, P. (2008). Counterfactual theories of causation. In *Stanford Encyclopedia of Philosophy*. Stanford: The Metaphysics Research Lab.

Menzies, P. (2009). Platitudes and counterexamples. In Beebee et al. [2009], (pp. 341–367).

Menzies, P. (2010). Norms, causes, and alternative possibilities. *Behavioral and Brain Sciences*, *33*(14), 346–347.

Menzies, P. (2011). The role of counterfactual dependence in causal judgements. In C. Hoerl, T. McCormack, & S. R. Beck (Eds.) *Understanding Counterfactuals, Understanding Causation*. Oxford University Press.

Menzies, P. (2012). The causal structure of mechanisms. *Studies in History and Philosophy of Science Part C*, *43*(4), 796–805.

Menzies, P. (2014). Counterfactual theories of causation. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Stanford: The Metaphysics Research Lab, spring 2014 ed.

Menzies, P., & List, C. (2010). The causal autonomy of the special sciences. In C. McDonald, & G. McDonald (Eds.) *Emergence in Mind*. Oxford: Oxford University Press.

Menzies, P., & Price, H. (1993). Causation as a secondary quality. *British Journal for the Philosophy of Science*, *44*(2), 187–203.

Mill, J. S. (1882). *A System Of Logic*. Harper and Brothers, Franklin Square, eighth ed.

Mumford, S. (2004). *Laws in Nature*. Routledge.

Noordhof, P. (1999). Probabilistic causation, preemption and counterfactuals. *Mind*, *108*(429), 95–125.

Northcott, R. (2007). Causation and contrast classes. *Philosophical Studies*, *139*(1), 111–123.

Papineau, D. (2013). Causation is macroscopic but not irreducible. In E. J. Lowe, & S. Gibb (Eds.) *The Ontology of Mental Causation*, (pp. 126–152). Oxford University Press.

Partee, B. H. (1977). Possible worlds semantics and linguistic theory. *The Monist*, *60*(3), 303–326.

Paul, L. A. (1998a). Keeping track of the time: Emending the counterfactual analysis of causation. *Analysis*, *58*(3), 191–198.

Paul, L. A. (1998b). Problems with late preemption. *Analysis*, *58*(1), 48–53.

Paul, L. A. (2000). Aspect causation. *Journal of Philosophy*, *97*(4), 235–256.

Paul, L. A. (2004). The context of essence. *Australasian Journal of Philosophy*, *82*(1), 170–184.

Paul, L. A. (2007). Constitutive overdetermination. In J. K. Campbell, M. O'Rourke, & H. S. Silverstein (Eds.) *Causation and Explanation*, chap. 13, (pp. 265–290). MIT Press.

Paul, L. A. (2009). Counterfactual theories. In Beebee et al. [2009].

Paul, L. A. (2012). Metaphysics as modeling: The handmaiden's tale. *Philosophical Studies*, *160*(1), 1–29.

Paul, L. A., & Hall, N. (2013). *Causation: A User's Guide*. Oxford University Press.

Peacocke, C. (1979). Deviant causal chains. *Midwest Studies In Philosophy*, *4*(1), 123–155.

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

Poli, R., & Seibt, J. (2010). *Theory and Applications of Ontology: Philosophical Perspectives*. Springer Verlag.

Price, H., & Weslake, B. (2009). The time-asymmetry of causation. In Beebee et al. [2009].

Psillos, S. (2009). Regularity theories. In Beebee et al. [2009].

Putnam, H. (1982). Why there isn't a ready-made world. *Synthese*, *51*(2), 205–228.

Quine, W. V. (1950). Identity, ostension, and hypostasis. *Journal of Philosophy*, *47*(22), 621–633.

Quine, W. V. (1961). On what there is. In T. Crane, & K. Farkas (Eds.) *From a Logical Point of View*, vol. 2, (pp. 21–38). Harvard University Press.

Quine, W. V. (1985). Events and reification. In E. Lepore, & B. McLaughlin (Eds.) *Actions and Events: Perspectives on the Philosophy of Davidson*, (pp. 162–71). Blackwell.

Ramachandran, M. (1997). A counterfactual analysis of causation. *Mind*, *106*(422), 263–277.

Ramachandran, M. (2004). A counterfactual analysis of indeterministic causation. In Collins et al. [2004], (pp. 387–402).

Ramsay, F. (1978). *Foundations*. Routledge and Kegan Paul, London.

Roberts, C. (2004). Context in dynamic interpretation. *Handbook of contemporary pragmatic theory*, (pp. 197–220).

Rodriguez-Pereyra, G. (1998). Mellor's facts and chances of causation. *Analysis*, *58*(3), 175Äì181.

Roxborough, C., & Cumby, J. (2009). Folk psychological concepts: Causation. *Philosophical Psychology*, *22*(2), 205–213.

Ruben, D.-H. (1994). A counterfactual theory of causal explanation. *Noûs*, *28*(4), 465–481.

Sartorio, C. (2004). How to be responsible for something without causing it. *Philosophical Perspectives*, *18*(1), 315–336.

Sartorio, C. (2005a). Causes as difference-makers. *Philosophical Studies*, *123*(1-2), 71–96.

Sartorio, C. (2005b). A new asymmetry between actions and omissions. *Noûs*, *39*(3), 460–482.

Sartorio, C. (2006). Disjunctive causes. *Journal of Philosophy*, *103*(10), 521–538.

Sartorio, C. (2010). The prince of wales problem for counterfactual theories of causation. In A. Hazlett (Ed.) *New Waves in Metaphysics*, (pp. 259–276). Palgrave McMillan, New York.

Schaffer, J. (2000a). Causation by disconnection. *Philosophy of Science*, *67*(2), 285–300.

Schaffer, J. (2000b). Trumping preemption. *Journal of Philosophy*, *97*(4), 165–181.

Schaffer, J. (2001a). Causation, influence, and effluence. *Analysis*, *61*(1), 11–19.

Schaffer, J. (2001b). Causes as probability raisers of processes. *Journal of Philosophy*, *98*(2), 75–92.

Schaffer, J. (2001c). Review of dowe's physical causation. *British Journal for the Philosophy of Science*, *52*(4), 809–813.

Schaffer, J. (2003a). Overdetermining causes. *Philosophical Studies*, *114*(1-2), 23–45.

Schaffer, J. (2003b). Overdetermining causes. *Philosophical Studies*, *114*(1-2), 23 – 45.

Schaffer, J. (2004a). Causes need not be physically connected to their effects: The case for negative causation. In C. R. Hitchcock (Ed.) *Contemporary Debates in Philosophy of Science*, (pp. 197–216). Basil Blackwell.

Schaffer, J. (2004b). Counterfactuals, causal independence and conceptual circularity. *Analysis*, *64*(4), 299–308.

Schaffer, J. (2005). Contrastive causation. *Philosophical Review*, *114*(3), 327–358.

Schaffer, J. (2007). Deterministic chance? *British Journal for the Philosophy of Science*, *58*(2), 113–140.

Schaffer, J. (2008). The metaphysics of causation. In *Stanford Encyclopedia of Philosophy*. Stanford: The Metaphysics Research Lab.

Schaffer, J. (2011). Contrastive causation in the law. *Legal Theory*, *16*(04), 259–297.

Schaffer, J. (2012a). Causal contextualisms. In M. Blaauw (Ed.) *Contrastivism in Philosophy: New Perspectives*. Routledge.

Schaffer, J. (2012b). Disconnection and responsibility. *Legal Theory*, *18*(Special Issue 04), 399–435.

Schaffer, J. (2014). The metaphysics of causation. In E. N. Zalta (Ed.) *The Stanford Encyclopedia of Philosophy*. Stanford: The Metaphysics Research Lab, summer 2014 ed.

Schlosser, M. E. (2007). Basic deviance reconsidered. *Analysis*, *67*(295), 186–194.

Schmitt, F. F. (1983). Events. *Erkenntnis*, *20*(3), 281–293.

Sider, T. (2001). *Four Dimensionalism: An Ontology of Persistence and Time*. Oxford University Press.

Sinnott-Armstrong, W. (2012). Free contrastivism. In M. Blaauw (Ed.) *Contrastivism in Philosophy: New Perspectives*. Routledge.

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

Stalnaker, R. C. (1968). A theory of conditionals. In N. Rescher (Ed.) *Studies in Logical Theory*, (pp. 98–112). Blackwell.

Stanley, J. (2000). Context and logical form. *Linguistics and Philosophy*, *23*(4), 391–434.

Steglich-Petersen, A. (2012). Against the contrastive account of singular causation. *British Journal for the Philosophy of Science*, *63*(1), 115–143.

Stone, J. (2009). Trumping the causal influence account of causation. *Philosophical Studies*, *142*(2), 153–160.

Strawson, P. F. (1992). Causation and explanation. In *Analysis and Metaphysics: An Introduction to Philosophy*. Oxford University Press.

Strevens, M. (2008). Comments on woodward, making things happen. *Philosophy and Phenomenological Research*, *77*(1), 171–192.

Swanson, E. (2010). Lessons from the context sensitivity of causal talk. *Journal of Philosophy*, *107*(5), 221–242.

Tännsjö, T. (2009). On deviant causal chains – no need for a general criterion. *Analysis*, *69*(3), 469–473.

Thomson, J. J. (2003). Causation: Omissions. *Philosophy and Phenomenological Research*, *66*(1), 81–103.

Unger, P. (1977). The uniqueness in causation. *American Philosophical Quarterly*, *14*(3), 177 – 188.

Varzi, A. C. (2007). Omissions and causal explanations. *Agency and Causation in the Human Sciences*, (pp. 155–167).

Weslake, B. (2013a). The problem of disjunctive explanations. *http://bweslake.s3.amazonaws.com/research/papers/weslake_disjunctive_explanations.pdf*.

Weslake, B. (2013b). Proportionality, contrast and explanation. *Australasian Journal of Philosophy*, *91*(4), 785–797.

Weslake, B. (forthcoming). A partial theory of actual causation. *British Journal for the Philosophy of Science*.

Wilson, J. M. (2006). Causality. In J. Pfeifer, & S. Sarkar (Eds.) *The Philosophy of Science: An Encyclopedia*, (pp. 90–100). Routledge.

Woodward, J. (2003). *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

Woodward, J. (2009). Agency and interventionist theories. In Beebee et al. [2009].

Yablo, S. (1992). Mental causation. *Philosophical Review*, *101*(2), 245–280.

Yablo, S. (2004). Advertisement for a sketch of an outline of a proto-theory of causation. In Collins et al. [2004], (pp. 119–137).