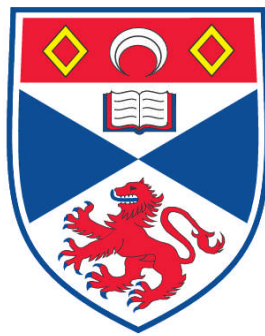# LEWIS, COUNTERFACTUAL ANALYSES OF CAUSATION, AND PRE-EMPTION CASES

## David Landsberg

## A Thesis Submitted for the Degree of PhD
## at the
## University of St. Andrews

## 2009

**Thesis Submission Form**

I, <u>David Landsberg</u>, hereby certify that this thesis, which is approximately <u>60,000</u> words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in <u>September 2005</u> and as a candidate for the degree of <u>PHD</u> in <u>January 2006</u>; the higher study for which this is a record was carried out in the
University of St Andrews between 2005 and 2009.

Date 06/08/09          Signature of Candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of <u>PHD</u> in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date 13/8/09          Signature of Supervisor

In submitting this thesis to the University of St Andrews we understand that we are giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. We also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. We have obtained any third-party copyright permissions that may be required in order to allow such access and migration, or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to Printed copy and electronic publication of thesis through the University of St Andrews.

Date 06/08/09          Signature of candidate

                       Signature of supervisor

**TITLE PAGE**

*Title***: Lewis, Counterfactual Analyses of Deterministic Causation, and Pre-emption Cases.**

*Author***: David Landsberg**

*Abstract*: Over the past few decades analyses of causation have proliferated in almost immeasurable abundance, and with two things in common; firstly, they make much of counterfactual dependence, and secondly, none of them successfully handle all the pre-emption cases. In this thesis, I fore-mostly investigate David Lewis' promising counterfactual analyses of causation (along with many others), and provide an extensive examination of pre-emption cases. I also offer my own counterfactual analysis of causation, which I argue can handle the problematic pre-emption cases, and therein succeed where so many other prominent analyses of causation have failed. I then conclude with some morals for the continuing debate.

0

**Lewis, Counterfactual Analyses of Deterministic Causation, and Pre-emption Cases.**

**David Landsberg**

Table of Contents

Index of core examples:

## Chapter 0: Introduction

## 0.1 Outline

Since Aristotle,[1] philosophers have inquired into the nature of causation. Why the interest? Moreover, why has the interest spanned millennia? An answer to the first question is that the concept of causation seems fundamental to our conception of the world, with common and scientific practice giving it a central role in the abundance of explanations about the workings of the known universe.[2] An answer to the second question is simply that a successful analysis of the concept of causation is yet to be had. This all suggests an embarrassment; we are plainly certain about an enormous amount of cases, no matter how mundane, concerning what causes what, but we are not in the least bit certain about what causation actually is! This seems perverse; for a complete understanding of nature surely requires a deep understanding of the concepts employed to understand nature in the first place. The question of what causation is then becomes a significant one.[3]

---

[1] See Aristotle, *Physics,* in Bostock [1999]

[2] Note that the terms 'causation,' 'causality,' 'cause,' and 'causal' feature in over 750,000 scientific articles published in an academic capacity since the turn of the century alone, as a refined Google-Scholar search will testify! This might suggest that such terms have some weight in scientific understanding. Whether the term *should* have such weight is of course a different question, with eliminativists such as Russell arguing that science does not need the concept at all (see 'on the notion of cause' in Slater, p.193-210, [1992]).

[3] Similar things may well be said about many other concepts investigated in metaphysics, indeed in philosophy.

I have just said that nobody is yet entirely clear about what causation is. What is a lot clearer is what a complete analysis of causation should provide us with, and that seems something like the following (see Schaffer [2007]):

First, it must provide us with an account of what the causal *relation* is, including an account of (i) what it is for two things to be causally connected (i.e. causal *connection*; might it be reduced to regularities (see Mackie [1974]), counterfactuals (see Lewis [1973b]), something else, or nothing at all (see Tooley [1987])?), (ii) what, regarding two causally connected things, it is for one to stand as a causal condition and the other as effect (i.e. causal *direction*; is it facts about time (see Hume [1748]), probabilistic distributions (see Papineau [1985]), a brand of perspectivalism (see Price [1996]), the asymmetry of over-determination (see Lewis [1979]), or something else?), and (iii) what it is for a thing to be a cause, as opposed to a mere causal condition or background factor for an effect (i.e. causal *selection*; is there no serious distinction to be had (see Lewis [1986a]), is it grounded in the activities of agents (see Hart and Honour [1959]), determined contrastively (see Schaffer [2005]), or something else?).

Second, it must provide us with an account of what the causal *relata* are, and therein supply us with an account of (i) the general sort of things they are (i.e. their *category*; are they events (see Davidson [1980] and Lewis [1986d]), facts (see Mellor [1995]), tropes (see Campbell [1990]), states of affairs (see Armstrong [1997]), features (see Dretske [1977]), situations (see Menzies [1989]), aspects (see Paul [2000]), or something else?), (ii) how fine-grained they are (i.e. their *individuation*; are they fine or coarse grained entities? (see Quine [1985], Kim [1976], and Bennett [1988] for some discussion)), and

(iii) how many there are (i.e. their *adicity*; is causation a two place (see Davidson [1993]) or a four place (see Schaffer [2005]) relation? Or something else?).

What features of causation will my thesis shed light on? I shall limit my investigation to *Lewisian style reductionist* (namely, *counterfactual*) analyses of the *deterministic* causal *relation,* and the issue of how they can be adapted to handle problematic cases of *pre-emption*. I say something about why I think this is a particularly important question for the causation debate here:

A first natural question is; why is an investigation into *reductionist* analyses of causation important? Causal reductionism is the thesis that causation can be reduced to/grounded in/determined by more basic and fundamental entities. There are different flavours of causal reductionism (depending on what those fundamental entities are taken to be), the one I am concerned with is the Lewisian package; i.e. to reduce causal facts to facts about counterfactuals, for facts about counterfactuals to be reduced to facts about laws of nature and possible worlds, and for facts about laws of nature and possible worlds to be reduced to facts about events. The package is attractive; it rids causation of mysterious entities (such as 'spooky' abstract objects that live in a platonic world beyond space and time[4]), and proposes to offer us a clear explanation of what causation is. All this stands in strong contrast to anti-reductionists and primitivists, who would either attempt to 'make sense' of causal facts in terms of an obscure ontology, or leave the nature of causation a persisting mystery. A successful reductionist analysis of causation is then a victory for

---

[4] In contrast, Lewisian concrete possible worlds are not in essence a mystery; we understand perfectly well what it is for something to be concrete (tables, chairs, protons, are all examples of concrete things).

one who seeks clear explanations, and furthermore for one who wishes to rid metaphysics of the questionable entities that often give the study a bad name.[5]

A second question is: Of reductionist analyses of causation, why investigate *counterfactual* analyses? The focus on counterfactual analyses has both a positive and a negative motivation. On the positive side, the notion of counterfactual dependence seems to have a lot to do with causation; paradigm cases of causation seem to be ones in which if the cause hadn't occurred the effect wouldn't have either, and the vast number of articles on counterfactual analyses of causation since Lewis' [1973b] influential paper is testament to this observation. On the negative side, the other popular reductionist strategy seems to have failed; we need but recall Lewis' influential denunciation of the regularity theory:

"It remains to be seen whether any regularity analysis can succeed in distinguishing genuine causes from effects, epiphenomena, and pre-empted potential causes- and whether it can succeed without falling victim to worse problems, without piling on the epicycles, and without departing from the fundamental idea that causation is an instantiation of regularities. I have no proof that regularity analyses are beyond repair… Suffice to say the prospects look dark. I think it is time to give up and try something else.

A promising alternative is not far to seek." (Lewis, p. 160, (1973b))

---

[5] See Schaffer [2008], for more about reductionist analyses of causation.

The promising alternative Lewis suggests is, as we might now suspect, to use conditionals of counterfactual dependence, i.e. conditionals of the form 'if C had not occurred, then E would not have occurred,' in some way to analyse the causal relation. It is this suggestion that will remain the central guiding thought in this thesis.

A third question is; why the general focus on *Lewisian* counterfactual analyses of causation? Lewis' analyses are by no means the only counterfactual analyses of causation in the literature.[6] However, there are four reasons why I (for the most part) shall focus on the works of Lewis. Firstly, insofar as Lewis is both responsible for the modern suggestion that causation be analysed in terms of counterfactuals and for the primary modern development of such theories, a strong focus on his works seems unavoidable. Secondly, Lewis' theories of causation have strongly influenced my own analysis (see chapter 4), and so an understanding of the strengths and failures of his theories brings us in a natural way to the presentation of my own. Thirdly, his theories about counterfactuals and causation fit together slickly (despite problems with both) and so an investigation into both seems natural; and fourthly, it would be nice to have a more or less concentrated resource on the adequacy of Lewis' analyses given their prominence in the literature and their importance for the ongoing debate; this thesis (amongst other things) will also attempt to provide this.

---

[6] See Noordhof [1999], Ramachandran [1997], Yablo [2004], and Woodward [2003] for alternative and prominent counterfactual approaches.

A fourth question is; why focus on *pre-emption* cases (general characterisations of which are given in section 3.0)? Lewis suggests in the above quote that a successful analysis of causation must be able to deal with the problems of effects, epiphenomena, and pre-emption, and do so without 'piling on' epicycles in a way that renders the analysis ad hoc or unduly complex. Of Lewis' requirements, I will spend most of the time investigating how counterfactual analyses handle cases of *pre-emption*. Cases of pre-emption are the most vicious problem for counterfactual analyses of causation; far more vicious than Lewis perhaps anticipated in his [1973b] paper. They are, perhaps, a problem of equivalent potential devastation to the project of analysing causation in terms of counterfactuals, as Gettier problems are to the project of analysing knowledge in terms of justified true belief.[7] The investigation into how a counterfactual analysis of causation can handle cases of pre-emption is then an urgent one (one I still think we can be optimistic about), with this thesis (to my knowledge) serving to be the most comprehensive cataloguing and examination of deterministic pre-emption cases to date.

A fifth question might be; why focus on *deterministic* causation, as opposed to *indeterministic* causation in this thesis? Unfortunately, for reasons of space, I have had to choose between whether to focus on deterministic as opposed to in-deterministic causation, with the choice between the two being more or less arbitrary. That said, Lewisian counterfactuals seem better disposed to be used to analyse deterministic

---

[7] See Gettier, [1963]

causation than indeterministic causation[8], and so the project of analysing deterministic causation in terms of such counterfactuals is more likely to succeed.

I proceed with the investigation as follows: In the rest of chapter 1, I set out the core piece of technical machinery appealed to in the rest of this thesis; counterfactual dependence between distinct actual events. First I explain Lewis' counterfactual relation (1.1), Lewis' causal relata (1.2), and then I briefly show how Lewis proposes to deal with the problems of effects and epiphenomena (1.3). This will set the stage for the investigation into Lewisian style counterfactual analyses of deterministic causation I wish to compare in chapter 2; Hume's naïve counterfactual analysis (see Hume, [1748]) (2.1), Lewis' counterfactual chains analysis (see Lewis, [1973b]) (2.2), Lewis' quasi-dependence analysis (see Lewis, [1986a]) (2.3), Lewis' influence analysis, (see Lewis, [2004a]) (2.4), [we shall also briefly visit Collins' would-be dependence analysis (see Collins, [2004]) in relation to the latter of Lewis' analyses in (2.5)]. This will in turn set the stage for the main part of this thesis, chapter 3, in which I investigate how these analyses handle the deterministic pre-emption cases. After disappointing results, I draw some morals from each of the analyses' failures and successes, which I in turn use to motivate my own analysis of deterministic causation based on counterfactuals, offered in chapter 4. By the end of the thesis, I hope I will have convinced the reader of two things; first, that none of Lewisian analyses of causation (nor their revised counterparts) are

_____

[8] See D. Edgington, 'Counterfactuals and the Benefit of hindsight' p. 12-28 in Dowe and Noordof (eds), for observations about the failure of Lewisian counterfactuals in contexts of indeterminism.

capable of handling all the problematic examples of deterministic pre-emption, and secondly, that my analysis is capable.

**0.2 Method**

So, the general task is to assess whether a counterfactual analysis of deterministic causation can be successfully developed, in particular with respect to handling the pre-emption cases. How should the task be approached? The usual method has been one of trial and error: We start with some intuitions about what causes what in some basic cases, present a counterfactual analysis of causation that proposes to accommodate such intuitions, and then show how the analysis handles the pre-emption cases; if it doesn't succeed, we add some extra revisions or epicycles to the analysis ('bells and whistles' as Lewis liked to call them) and then see if that works, or go back to the drawing board for yet another counterfactual analysis of causation. The main proponent of this sort of method (at least implicitly) has been Lewis himself.[9] How legitimate is this method?

For starters, we may very well question how legitimate it is to take intuitions about cases to play the role of evidence for what the concept of causation should look like. We might do so in two ways; firstly, we might question whether we should provide some sort of empirical evidence to support our intuitions about cases (drawn from the physics about the cases we are interested in analysing), in order to vindicate the analysis that draws

---

[9] See Lewis' [1973b], [1986c], and [2004a] for demonstrations of his use of the method of 'trial and error'.

from such intuitions. And secondly, we might question what to do when we have conflicting or confused intuitions about particular cases of causation.

In response to the first question, I would say that the criterion that all our intuitions about cases of causation must have empirical evidence in support of them is too strong- for there are possible cases of causation for which there can be no empirical evidence, but we must in turn account for[10] (namely, distant possible but non-actual cases of causation, such as the sort contained in some cases of trumping pre-emption). Nevertheless, we should adhere to the (weaker) criterion that our intuitions about cases of causation *should not conflict* with whatever empirical evidence there is about such cases (if indeed there is any at all), and proceed in that way. Luckily for us, the majority of cases of causation we shall be interested in are so uncontroversial that it would be superfluous to demonstrate how they are consistent with actual world physics (I talk namely of the BILLY AND SUZY and CAR RACE examples), and so the issue of controversy about using intuitions in this respect to make judgments about such cases does not even arise.

In response to the second question, I would respond that (as luck would have it) intuitions about what causes what in the pre-emption cases which we shall be visiting are pretty solid. Cases of trumping pre-emption and pre-emptive prevention are the only cases we

---

[10] Must we account for distant possible cases of causation? There are some who engage in the project of analysing causation as it occurs in all nomologically possible worlds, such as Dowe (see Dowe, Physical Causation, [2000]. Our ambitions will be a lot bolder- we shall be engaging in the project of analysing causation as it occurs in all logically possible worlds, and will therefore attempt to analyze cases which are beyond the realm of physical possibility.

shall discuss which have received strongly conflicting intuitions, but I shall offer a treatment of such cases later to resolve such issues.

Issues about the status of intuitions aside, the method of trial and error has a virtue and a vice in equal proportion, as follows:

Firstly, *the virtue*: The virtue of the method is that it's a quick and efficient way to assess whether any particular counterfactual analysis might work in getting cases of causation right, that is, in analysing all the causes as causes and non-causes as non-causes from case to case. Following this it then seems a good way to assess whether a proto-type analysis is both necessary and sufficient for causation, or at least is along the right lines.

Secondly, *the vice*: The vice is that by using this method, it's pretty easy to fall into an almost endless pit of revision making, often with the end result being an analysis of causation with so many epicycles, that it's either hard to understand or far removed from the basic intuition that causation is strongly linked to simple instantiations of regularities and counterfactual dependency in the first place, despite its relative success in dealing with cases. It is precisely the vice of having 'too many epicycles' which Lewis warns us to avoid in the earlier quote.

To draw out the thought, the method of trial and error seems a helpful way of guiding us towards an *extensional* analysis of causation (i.e. an analysis of causation that gets all the cases right), but not necessarily an effective one at finding a simple *intensional* analysis

of causation (i.e. an analysis of causation that also provides a simple and intuitive meaning of the term), something we also want. The thought suggests a moral for a successful theory of causation; it should have *strength* in dealing with all the cases correctly, but should also be the overall *simplest* in its means to achieving this end. In developing a successful analysis of causation, I propose that strength in dealing with cases always takes place over simplicity, and this will motivate the revisions I make to the five prominent analyses of causation I mentioned earlier. However, when two theories are equally strong, I propose the simplest of them should prevail, and this shall be reflected when I reject an early analysis of causation I once proposed (see appendix), in favour of the one I now offer (see chapter 4).

I shall use the method of trial and error in this thesis, with disappointing results for the Lewisian analyses. We shall not be left empty handed however, for putting these analyses through the gauntlet of trial and error and epicycle making will make apparent the virtues and vices of each of them, in turn suggesting some key morals for what a successful analysis of causation should look like. In chapter 4, I shall take these morals to motivate my own analysis of causation.

**Chapter 1: Counterfactuals**

**1.1 Lewis' Causal-Counterfactual Relation**

In 'Counterfactuals' [1973a], Lewis presents theories for the truth conditions of counterfactual conditionals. In this section I present a brief outline of one these theories. The investigation will serve to provide a framework for the basic 'building blocks' of the analyses of causation investigated in the next chapter. Lewis' theories about counterfactuals are certainly not the only ones in the literature,[11] but I appeal to Lewis firstly because he provides the most popular working theory about counterfactuals to date, and secondly because most of the analyses of causation we shall be investigating (including my own) themselves appeal to something like his theories. Following this, I shall give an explanation of what Lewis holds laws to be (1.1), which will in turn help explain Lewis' analysis of comparative similarity (1.2), which will in turn help explain the Lewisian notions of counterfactual dependence and implication (1.3).

**1.1.1   Laws of nature**

A Lewisian style theory of laws[12] may be presented as follows: Following the Mill, Ramsey, and Lewis systems approach;

---

[11] See Goodman [1947] and Kvart [1986] for some different theories about counterfactuals.

[12] Lewis, [1999], p. 41.

- Laws of nature are theorems of the axiomatisation of the facts that best balance explanatory simplicity and strength.

Laws may either be deterministic or indeterministic. First, determinism:

- (1$^{st}$ level) The thesis of determinism is true iff our world is deterministic.
- (2$^{nd}$ level) A world W is deterministic iff W's laws are deterministic.
- (3$^{rd}$ level) A system of laws is deterministic iff no two divergent worlds both conform to those same laws[13].

Determinism can be strong or weak, depending on how we take the term 'divergence' at the 3$^{rd}$ level, such that:

- A system of laws is strongly deterministic iff no two worlds can locally diverge and both conform to the same laws (two worlds locally diverge iff these worlds differ in *some* of their properties).
- A system of laws is weakly deterministic iff no two worlds can globally diverge and both conform to the same laws (two worlds globally diverge iff these worlds differ in *all* their properties).

Now, we are ready to express the thesis of indeterminism:

---

[13] Another formulation of determinism is that a set of laws is deterministic iff the total set of laws together with the total set of events at any one time is only consistent with one possible past and future.

- The thesis of indeterminism is true iff the thesis of determinism is false.

We note two things: Firstly, the analysis of causation in which we shall be interested is an analysis of deterministic causation. An analysis of deterministic causation is an analysis of causation in worlds in which nomological determinism is true. Secondly, we note that a theory of laws plays a role in how one of the central components in Lewisian counterfactual semantics - the notion of 'comparative similarity' - is to be understood.

## 1.1.2 Comparative similarity

Following Lewis, counterfactual conditionals are analysed in terms of a possible worlds semantics. A possible worlds semantics gives truth conditions for counterfactuals in terms of relations between possible worlds (whatever we take those possible worlds[14] to be). The relation appealed to for our analysis of causal counterfactuals is one of *comparative similarity* between possible worlds: W1 is said to be *more similiar* to W0 than W2 if W1 has more relevant properties in common with W0 than W2 does (Lewis takes this notion to be a reflexive, non-symmetric, non-transitive, and totally connected

---

[14] From here on in, if I talk about 'worlds' simpliciter, I am still talking about 'possible worlds'. We shall not be visiting 'impossible worlds' in this thesis. Lewis' modal realism takes possible worlds to be spatio-temporally discrete universes as real as the actual world; we shall not be imposing this view in this thesis. See Lewis, [1986e].

relation[15]). The relevant properties that feature are nomological or local, such that the notion of comparative similarity can be given truth conditions as follows:

W1 is more similar to W0 than W2 iff [16]

- As a matter of 1st priority, W1 avoids *widespread* violations of W0's nomological laws more than W2.

- As a matter of 2nd priority, W1 *maximises* the spatio-temporal perfect match of particular events from W0 more than W2.

- As a matter of 3rd priority, W1 avoids *local* violations of W0's nomological laws more than W2.

- As a matter of 4th priority, W1 secures *approximate* similarity of W0's particular events more than W2.

## 1.1.3 Truth conditions

Lewisian-style truth conditions for a conditional of counterfactual *implication* 'if A had occurred, then C would have occurred,' may be given as follows. Where A and C are events, '*o*' a predicate of occurrence, and '▢▢' the counterfactual operation, then:[17]

---

[15] Lewis, p. 164, [1973b]

[16] Lewis, p.47-8, [1986d]

- '*o*A □→ *o*C' is true iff i) there are no possible worlds in which A occurs, or ii) the most similar worlds to the actual world in which A occurs are worlds in which C occurs.[18]

I ignore counterfactuals which are true because the *first* condition is satisfied,[19] for if this is the case, then the counterfactual is trivially true. Consider the following (trivially true) counterfactual in this respect:

---

[17] Note that the following analysis of counterfactual implication can itself be analysed in terms of the comparative similarity relation we just discussed. Namely, the most similar worlds to the actual world in which A occurs are worlds in which C occurs iff there is some world W1 in which A and C occurs which is more similar to the actual world W0 than any world W3 in which A occurs and C doesn't, where it is not the case that there is a world W4 which is not W3 which in turn is a world in which A and C occurs and is also more similar to the actual world than W1.

[18] This Lewisian analysis differs from an alternative account he offers [1973a], as follows; where P and Q are propositions then, 'P □→ Q' is true iff a P and Q world is closer than any P and ~Q world. Lewis argues that the latter sort of truth conditions for counterfactuals is superior to the one I am appealing to in this thesis, given the latter ones do not assume that there is always a *most* similar class of possible worlds (i.e. the latter ones do not rely on the 'limit assumption.').

[19] In this thesis I shall be ignoring interesting cases in which our Lewisian semantics might have to be revised to include impossible worlds in order to capture some cases of causation correctly. Such interesting cases might include absences of impossibilia that cause things in non trivial ways, and necessary individuals that cause things in non trivial ways- but all this will have to be ignored for the sake of concentrating on pre-emption cases.

- If the presence of round squares at t had occurred, then Margaret Thatcher's joining of the Labour party/the omnipresence of purple hedgehogs at a time etc. would have occurred.

In this thesis, I shall only be interested in counterfactuals which are true because the *second* condition is satisfied, for if this is the case, then the counterfactual is non-trivially true. Consider the following (non trivially true) counterfactuals in this respect:

- If the bringing of my umbrella had occurred, my staying dry would have occurred (which is true just in case the most similar worlds to the actual world in which the bringing of my umbrella occurs are worlds in which my staying dry occurs)
- If my throwing of a ball had occurred, then the smashing of the window would have occurred (which is true just in case the most similar worlds to the actual world in which my throwing of a ball occurs are worlds in which the smashing of the window occurs).

The particular type of counterfactual which we shall be interested in is non-trivial counterfactual *dependence* between events. The truth conditions for a conditional of counterfactual dependence 'if A hadn't occurred, then C wouldn't have occurred' may thereby be given as follows:

- '~oA □□ ~oC' is true iff i) there are no possible worlds in which it is not the case that A occurs, or ii) the most similar worlds to the actual world in which it is not the case that A occurs are worlds in which it is not the case that C occurs.

Now we have developed the counterfactual notions we shall be interested in, it is time to consider some initial problems for them.

## 1.1.4 The future similarity problem

Without our Lewisian analysis of the similarity relation (and taking the similarity relation as an unanalysed primitive), Lewisian counterfactuals would be subject to what is popularly known as 'the future similarity problem'[20]. To see this problem, consider the following (now well worn) counterfactual:

- If Nixon's button pressing had occurred, then a nuclear holocaust would have occurred.

Suppose this counterfactual is true, i.e. suppose that Nixon had access to a button in the Pentagon which had the power to launch hundreds of nuclear warheads with the disposition to destroy much of the earth as we know it.[21] Now, on a view in which we take the similarity relation as an unanalysed primitive, the above counterfactual might

---

[20] See Kit Fine, [1975]

[21] Suppose also this all takes place under deterministic laws of nature.

24

seem to turn out (incorrectly) as false. This is because it might seem that the most similar possible worlds to the actual world in which Nixon's button pressing occurred, are *not* worlds in which a nuclear holocaust occurred, or in other words, that button pressing/non-holocaust worlds are more similar to the actual world than button pressing/holocaust worlds. In detail, the reason such worlds might seem more similar is that there are more one to one correspondences of instantiated local properties between those worlds and our world, than between worlds in which there are holocausts and our world- they simply don't have all the mess of a holocaust world! On such a view, it would then seem as if our Lewisian semantics, equipped with an unanalysed notion of similarity, would fail to correctly analyse the truth conditions for counterfactuals, such as the Nixon counterfactual above.

The 'future similarity problem' is precisely why Lewis gave the thorough analysis of the similarity relation we outlined in section 1.2. On such an analysis, we can say that in ordering worlds in respect of similarity to the actual world, it is a matter of *first priority* that worlds should in general preserve the actual world's laws of nature. On this approach, button pressing holocaust worlds are more similar to the actual world than button pressing no holocaust worlds, because worlds in which the former states occur preserve the actual world's laws of nature much more than worlds in which the latter states occur. Thus, according to the analysis of the similarity relation in 1.2, we can safely assume the truth of the above counterfactual; the correct result.[22]

---

[22] We ignore indeterministic variations of the future similarity problem, given we are ignoring indeterministic causation in this thesis.

## 1.2 Lewis' Causal-Counterfactual Relata

### 1.2.1 Events:

Before engaging in this thesis' major project, I should assume something about what the causal relata are taken to be, for the sake of argument. Here, I offer a somewhat dogmatic stipulation of a theory, as follows:

- The causal relata are distinct actual events.[23]

Much later I shall consider allowing omissions into the causal relata, but I shall not discuss this complication here. For now, some elaboration on the above theory:

- Events are instantiations of intrinsic properties local to spatio-temporal regions.[24]
- Omissions are the absence of intrinsic properties local to spatio-temporal regions.
- Events are distinct iff they share no common mereological part.[25]

---

[23] Given the coherency of closed causal loops, it might arguably be possible for an event to cause itself, and thus for a cause and effect not to be distinct, but for simplicity we shall ignore such cases unless stated otherwise, and keep our attention to everyday run of the mill cases of causation in which the causal relata are distinct. I shall also ignore the issue of how natural events have to be to play the role of causal relata, as this issue shall not bother us either.

[24] See Lewis, p. 244-249, [1986b].

- Events are fine-grained iff all their properties are essential to them (such that if the event had changed it would not have occurred).[26]

- Events are coarse-grained iff only some of their properties are essential to them (i.e. the event could still have occurred with somewhat different properties).[27]

- Properties are sets of possibilia.[28]

- Natural properties are ungerrymandered and ungruesomated properties.[29]

- Intrinsic properties are those properties that can never differ between duplicates.[30]

- Extrinsic properties are those properties that can differ between duplicates.

## 1.2.2 Fine and coarse grained events

Some extra details about the difference between coarse and fine-grained events will come in handy later. The difference may be explained as follows:

[25] See Lewis, p. 212, [1986b]. Lewis writes "…two events are distinct if they have nothing in common: they are not identical, neither is a proper part of the other, nor do they have a common part."

[26] I take the term 'fine-grained events' to be synonymous with 'concrete events,' and 'fragile events,' and 'low-order events'.

[27] I take the term 'coarse-grained events' to be synonymous with 'robust events,' 'high-order events'.

[28] Lewis, p.50, [1986e].

[29] What it is for a property to be ungerrymandered and ungruesomated is a topic in its own right. We don't have time to adequately analyse a theory of natural events properly here; it suffices to say that if a type of event is well entrenched in our discursive practices then that type of event stands as a good candidate for being natural. So events like football games, sunrises, explosions, and planetary orbits are all natural in this way.

[30] Lewis, 'defining intrinsic,' p. 121, in his [1999]

- If A is a fine-grained event, it is an equivalence class of events with one member; the class has no other members that are distinct from that member, and the precise intrinsic manner in which A occurs is the only way A could have occurred.

- If A is a coarse-grained event, it is an equivalence class of fine-grained & non-compossible events {A1, A2… etc.}, such that A could have occurred in different ways.

The occurrence conditions for events (in terms of their equivalence classes) can then be given as follows:

- An event (either coarse or fine-grained) occurs iff one of the members of its equivalence class occurs.
- An event (either coarse or fine-grained) does not occur iff none of the members of its equivalence class occurs.

Coarse-grained events locally supervene on fine-grained events. That is, there is no change in a coarse-grained event without a change in a fine-grained event, and if a fine-grained event occurs then it necessitates the occurrence of a coarse-grained event. Furthermore, if a coarse-grained event A occurs, it occurs because one of the fine-grained members of its equivalence class occurs, in which case we say the fine-grained member is a 'realiser' of the coarse-grained event.

## 1.2.3 Distinct events

Some more preliminaries: Shortly I shall suppose that counterfactual dependence between *distinct* events is sufficient for causation. Counterfactual dependence between events simpliciter is *not*. To demonstrate this, we reduce to absurdity the proposition that (1) counterfactual dependence between events simpliciter is sufficient for causation, as follows:

An example due to Collins, Hall, and Paul:[31] Assume that my shutting of the door does not cause me to slam it (as it wouldn't), and that I shut the door by slamming it. Now, it is true that *if the shutting had not occurred, then the slamming would not have occurred either*, but this does not imply a causal relation, and so (1) must be false.

An example due to Kim:[32] Assume that my writing 'r' twice in succession does not cause me to write 'Larry' (as it wouldn't), and that I write 'Larry'. Now, it is true that *if I hadn't written 'r' twice in succession, then I would not have written 'Larry'*, but this does not imply a causal relation, and so (1) must be false.

---

[31] Collins, Hall, & Paul, p.21, [2002]

[32] Kim, p. 206, [1975]

An example due to Collins, Hall and Paul:[33] Assume a certain shattering of a glass consists of all its shards flying different directions, and that the shattering of a glass did not cause its shards to fly in different directions (as it wouldn't). Now, suppose a glass, is shattered, it is true that *if the glass had not shattered, then the shards would not have flown in different directions*, but this does not imply a causal relation, and so (1) must be false.

An example due to Coady:[34] Assume that the First World War did not cause its battles to occur (as it wouldn't), and that WWI consists in part of many local battles; the Somme, of the Reign etc. Now it is true that *if World War I had not occurred, then none of its battles would have occurred*, but this does not imply a causal relation, and so (1) must be false.

Our initial assumptions are true; there is no causation going on in these cases. And they are true for the following reasons: In the first pair of cases, the relata stand in an inappropriate logical relationship; shutting the door is logically necessary for its slamming, and writing 'r' twice is logically necessary for writing 'Larry'. In the second pair of cases, the relata stand in an inappropriate mereological relationship; the battles were mereological constituents of WWI, and the flying shards of glass were mereological constituents of the shattering. In order to exclude these sorts of cases from implying

---

[33] Collins, Hall, & Paul, p.21, [2001]

[34] Coady, [2004]

causal relationships, we deny (1), that counterfactual dependence between indistinct events is sufficient for causation.

## 1.3 The problems of effects and epiphenomena

## 1.3.1 Backtrackers

These preliminaries aside, we are now ready to state a sufficiency condition for causal conditions in terms of counterfactuals:

- (SC) If C and E are distinct actual events, and if C had not occurred then E would not have occurred, then C is a cause of E.

Lewis offers this sufficiency condition (SC) in 'On the Plurality of Worlds' (Lewis, p.26 [1986]). Does (SC) hold upon Lewisian counterfactual semantics? There are two types of cases in which we might argue that it doesn't.

The first type of case is one in which SC purportedly fails to respect the difference between 'true causes' and 'background causal conditions.' To see the prospects for this distinction, consider a case in which I strike a match in an oxygenated environment, and the match lights as a result. Two counterfactuals are true; i) if my striking of the match hadn't occurred, then the match lighting wouldn't have occurred, and ii) if the presence of oxygen hadn't occurred then the match lighting wouldn't have occurred. SC then tells

us that both the presence of oxygen and the striking of the match are each a cause of the match lighting, where, it may be argued, this is false; only the striking of the match is a true cause of the match lighting; the presence of oxygen is a mere background causal condition. How should we respond to this claim? We can save SC in the face of these sorts of cases by saying that what we denote by 'cause' in the analysis is really something like a causal condition, where the term 'causal condition' is inclusive enough to include both true causes along with background causal conditions; and thereby  say that SC respects whatever distinctions there might be about different types of causes. The appendix  will have more to say about this distinction, but we must ignore it for now.

The second (more threatening) type of case is one in which SC purportedly generates 'spurious backwards causes' (i.e. spurious cases in which something is said to cause something else in its past), because it allows some 'backtracking' counterfactuals to enter into the analysis (i.e. a counterfactual in which the antecedent is in the causal future of the consequent). An example of a backtracking counterfactual is the counterfactual 'if I had not come to the seminar today, then I would not have started philosophy 7 years ago.' Regardless of whether we think backtracking counterfactuals such as this one should turn out true or not, it is clearly *not* sufficient for the causal claim that my coming to the seminar today caused my starting philosophy 7 years ago. Thus, to confirm whether SC is true upon Lewisian counterfactual semantics, 'backtracking counterfactuals' must be ruled out from being entertained by SC. To see in detail how Lewis proposes this is done, I will discuss the problem of 'effects' and 'epiphenomena'.

## 1.3.2 The problem of effects

Consider the diagram below; C causes E, but E does not cause C (grey circles represent events that actually occur, arrows represent causal connections and their causal direction, and the temporal order runs from left to right):

C ⚫ ⟶ ⚫ E
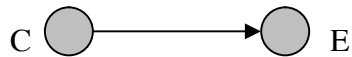
An example: The throwing of the ball (C) caused the window to shatter (E), but the window shattering did not cause the throwing of the ball. Now consider the two following counterfactuals:

1. If the throwing of the ball (C) had not occurred, then the window shattering (E) would not have occurred.
2. If the window shattering (E) had not occurred, then the throwing of the ball (C) would not have occurred.

The first counterfactual is clearly sufficient for a causal relationship; the ball is a cause of the smashing of the window, however the second counterfactual is not; the smashing of the window is not a cause of the throwing of the ball. The problem confronting SC is then how Lewisian counterfactual semantics rules out the second counterfactual from being sufficient for a causal relationship.

### 1.3.3 The problem of epiphenomena:

Two events E and F are a group of epiphenomena iff E and F do not cause each other but have a common cause. Consider the diagram below; C causes E & F, E does not cause C or F, & F does not cause C or E, such that E and F are a group of epiphenomena.

E

C

F
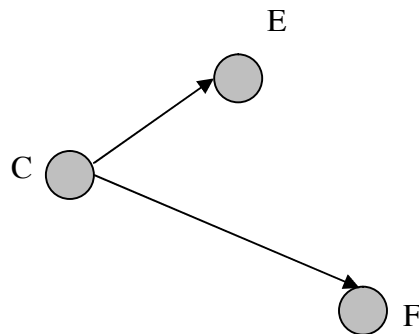
An example: The fall in pressure (C) caused the meter on the barometer to rise (E) and the storm (F), but the barometer rising (E) did not cause the storm (F) (as such, the

barometer rising (E) and the storm (F) together constitute a group of epiphenomena). Now consider the following counterfactual:

3. If the barometer rising (E) had not occurred, then the storm (F) would not have occurred.

This counterfactual is clearly not sufficient for a causal relationship; the barometer rising (E) is not a cause of the storm (F). The problem confronting SC is then how Lewisian counterfactual semantics rules out this counterfactual from being sufficient for a causal relationship.

The cases of effects and epiphenomena have been set up, and we are now ready to pose our problematic question. How do Lewisian counterfactual semantics rule out backtracking counterfactuals (like 2 and 3) from being true?

**1.3.4 Lewis' proposed solutions**

One way to rule out backtrackers from entering into SC might be to state some condition to the effect that *causes must temporally precede their effects*; in which case the second counterfactual (2) would be prevented from generating a spurious claim of causal dependence. Lewis argues that we should deny this response for three reasons:

- It fails to solve the closely related problem of epiphenomena, because

epiphenomenon E *does* precede its spurious effect F.

- It rules out the possibility of backward or simultaneous causation *a priori*.

- We would no longer be able to analyse the direction of time in terms of the direction of causation on pain of circularity.

After quickly rejecting a temporal theory as a solution to the problems of effects and epiphenomena, Lewis proposes his own solution as follows:

"The proper solution to both problems, I think, is flatly to deny the counterfactuals that cause the trouble. If E had been absent [in either case], it is not that C would have been absent (and F with it, in the second case). Rather, C would have occurred just as it did but would have failed to cause E. It is less of a departure from actuality to get rid of E by holding C fixed and giving up some or other of the laws and circumstances in virtue of which C could not have failed to cause E, rather than to hold those laws and circumstances, fixed and get rid of E by going back and abolishing its cause E."[35]

So why does Lewis think that not-E & not-C occurring worlds violate the laws and supposed circumstances more than not-E & C occurring worlds in the two examples we just visited? To understand the Lewisian response, a quick re-cap on some definitions of *determinism* and *over-determination* is necessary. The thesis of *determinism* holds that the collection of events at any one time lawfully determines a unique past and future,

---

[35] Lewis, p.11, [1973b]

such that any event is pre-determined throughout the past and post-determined throughout the future. If an event has only one lawful determiner at a time, then it is lawfully determined, but not lawfully over-determined at that time, but if it has more than one lawful determiner at a time, then it is lawfully *over-determined* at that time.

Lewis' response is as follows: Lewis observes that, as a matter of contingent fact, it is usually the case that an event is vastly more *over-determined* by events in its causal future than in its causal past. Furthermore, this fact has consequences for which counterfactuals turn out true, in particular, the ones that feature in our examples in the problems of effects and epiphenomena:

In the example for the problem of effects, the spurious causal-counterfactual (2) 'if the shattering (E) had not occurred, then the throw (C) would not have occurred' is false. Rather, if the shattering (E) had not occurred, then the throw (C) would still have occurred but would have failed to cause the shattering (E). This is because (according to Lewis) a world in which the shattering doesn't occur but the throw still does is more similar to the actual world than any worlds in which neither the shattering nor the throw occurs. This is in turn because the former sort of world violates far fewer of the actual laws of nature than the latter, given the latter sort of worlds would have to violate laws which render the throw (C) as grossly *overdetermined* by events in its *causal future* (such as the wind's motion after the throw, the particular fatigue in the thrower's arm, other people's memories of the throw etc.), where the former sorts of world do not.

In the example for the problem of epiphenomena, the spurious causal-counterfactual (3) 'if the barometer rising had not occurred, then the storm would not have occurred' is false. Rather, had the barometer level rising not occurred, then the storm would still have occurred and caused the fall in pressure but would not have caused the barometer level rising. This is because (according to Lewis) a world in which the barometer rising still occurs but the storm doesn't, is more similar to the actual world than any worlds in which neither the barometer rises nor the storm occurs. This is in turn because the former sort of world violates far less of the actual laws of nature than the latter- the latter sort of worlds would have to violate laws which render the storm as grossly *over-determined* by events in its causal future (such as all the messy aftermath of the storm!), where the former sorts of world do not.

The upshot of Lewis' response is that due to facts about the asymmetry of over-determination, it is more similar to the actual world to get rid of an event and fix all of its ancestral causes, than to get rid of an event and also get rid of its ancestral causes, and thus backtrackers turn out false upon Lewisian semantics. However, as Lewis himself notes:

> "The asymmetry of overdetermination is a contingent, de facto matter. Moreover, it may be a local matter, holding near here but not in remote parts of space and time. If so, then all that rests on it - the asymmetries of counterfactual

38

dependence, of causation and openness - may likewise be local and subject to exceptions."[36]

In short, the problem for Lewis' proposed solution is that it doesn't offer a complete solution to the problem concerning the ruling out of backtrackers from his counterfactuals, given there may be possible worlds, or areas of this world itself, in which there is no such asymmetry of overdetermination of the required sort to provide us with the proper causal-counterfactual asymmetries. We can see the incompleteness of Lewis' proposed solution by considering the following counterfactual:

- If the sum of all the events (after time t) (A) hadn't occurred, my sitting on a chair (at t-1) (C) wouldn't have occurred

In this backtracking counterfactual, there is no 'asymmetry of over-determination' which rules it out from being true upon Lewisian semantics. All the future over-determiners are contained within the event specified in the antecedent, such that the most similar worlds in which the sum of all the events after t don't occur are worlds in which I don't sit on a chair, given that these worlds would violate fewer of the deterministic laws than worlds in which the sum of all the events after t don't occur but in which I still sit on a chair. Accordingly, SC incorrectly analyses A as a cause of E. How should we respond to this problem?

---

[36] Lewis, p. 50-51, [1979]

The problem is that we can no longer use Lewisian counterfactuals to analyse the causal asymmetry, and this suggests that we either revise our counterfactual semantics, or look elsewhere for an analysis of the causal asymmetry itself. I shall not investigate any of these projects here, as such an investigation would lead us too far afield. It suffices to say that a good response to the problem of ruling out backtrackers might be found in analysing the causal asymmetry in terms of some other asymmetry not accounted for by Lewis' asymmetry of over-determination, and that the counterfactual strategy is thereby not doomed from the start. For the sake of argument, we will have to proceed with some working theory for the causal asymmetry, and so we will impose the condition that if C causes E then C precedes E (i.e. the theory that Lewis rejected to begin with!). Given we shall not be treating any cases of backwards or simultaneous causation in this thesis, or be worried about analysing the direction of time in terms of the direction of causation, this will not be problematic for us.

One may well ask here, if Lewisian counterfactuals are no good at analysing causal *asymmetries*, what good are they for analysing causation? As we shall see later on, in spite of issues of causal asymmetry, there is still good hope that such counterfactuals might be good at analysing a notion better fitted to that of causal *connection* (where we say C and E are causally connected iff one causes the other). It is with this suggestion that I proceed.

**Chapter 2: Counterfactual Analyses of Causation**

In this chapter I will explain the following five prominent Lewisian style counterfactual analyses of deterministic causation, in light of the previous investigation into Lewisian counterfactuals:

1. Hume's naïve counterfactual analysis (see Hume, [1748])

2. Lewis' counterfactual chains analysis (see Lewis, [1973b])

3. Lewis' quasi-dependence analysis (see Lewis, [1986a])

4. Lewis' influence analysis (see Lewis, [2004a])

5. Collins' would-be dependence analysis (see Collins, [2004])

I will then move on to how they handle the pre-emption cases (chapter 3).

**2.1 Hume's counterfactual analysis [1748]**

Hume wrote:

> "We may define a cause to be an object followed by another, and where all objects, similar to the first, are followed by objects similar to the second. Or in other words, where, if the first object had not been, the second never had existed."[37]

---

[37] Hume, [1748]

As Lewis observed,[38] the first sentence proposes a regularity theory of causation, which is different from the view espoused by the second sentence, which proposes a counterfactual analysis of causation. The two theories may be expressed more explicitly as follows (leaving issues to do with the relata aside for the moment):

Hume's regularity theory:

- C causes E iff objects similar to C are regularly followed by objects similar to E.

Hume's basic counterfactual theory:

- C causes E iff it's the case that if C had not been, then E wouldn't have been, and C precedes E.

The former theory is of the sort of regularity theory that Lewis rejects.[39] Arguments for its falsity are commonplace - it's pretty easy to see why it seems neither necessary nor sufficient for causation:

For starters, the analysis seems not *necessary* for causation because it seems possible for C to cause E, without there being a regular connection between similar Cs and Es. An

---

[38] Lewis, p. 160, [1973b]

[39] Ibid.

example; supposing a world in which the big bang causes the expansion of the cosmos, we may presume that there certainly are not any similar objects in the world similar to the big bang that are followed by objects similar to the expansion of the cosmos, with any plausible degree of regularity.

The regularity theory also seems in*sufficient* for causation because it seems possible for it to be the case that C doesn't cause E, although objects similar to C are regularly followed by objects similar to E. An example; we may suppose that a tsunami off the Japanese coast does not cause the fluttering a butterfly's wings in South America, but that all objects similar to the tsunami are regularly followed by objects similar to the fluttering of a butterfly's wings.

Perhaps problems like these can be got around, namely by tweaking the analysis into the form advocated by its most recent proponents (see Mackie [1974]) but like Lewis, I shall not engage in the project of trying to do this. Suffice to say that it is Hume's suggestion that causation can be analysed in terms of counterfactual dependence that is the key suggestion that provides a *historically* grounded motivation for the thesis that causation has something to do with counterfactuals.

Given our Lewisian counterfactual semantics and a theory for the counterfactual relata, we can restate Hume's theory as follows:

- C causes E iff if C hadn't occurred E wouldn't have occurred (i.e. the most similar worlds in which C doesn't occur are worlds in which E doesn't occur), C precedes E, and C and E are distinct actual events.

How does the analysis deal with a paradigm case of causation? Suppose I throw a ball, and the window shatters. My throw (C) causes the window shattering (E) as follows: if my throw (C) hadn't occurred, the window shattering (E) wouldn't have occurred (i.e. the most similar worlds in which my throw (C) doesn't occur, are worlds in which the window shattering (E) does not occur), my throw (C) precedes the window shattering (E), and both are distinct actual events. So far so good.

## 2.2 Lewis' counterfactual chains analysis [1973]

Lewis' [1973b] analysis of causation differs slightly from Hume's naïve counterfactual analysis. Lewis wants to analyse causation in terms of chains of counterfactual dependence (rendering causation a transitive relation) in order to analyse basic cases of early pre-emption correctly (cases we shall come to later). Note that counterfactual dependence is not itself a transitive relation (i.e. if ~oA □→ ~oC, and ~oC □→ ~oD, then it does not follow that ~oA □→ ~oD). To give an analysis of a transitive causal relation in terms of counterfactual dependence Lewis gives the truth conditions for causation in terms of counterfactual *chains*, where the truth conditions for counterfactual chains can be expressed as follows:

- There is a chain of counterfactual dependence from E to C iff there are some actual events (E, D, D1 etc… C), such that E counterfactually depends on D, D on D1 etc…. C.

Lewis' 1973 analysis of causation can then be expressed as follows; where C and E are distinct actual events, then:

- C causes E iff there is a counterfactual chain of dependence from E to C.

How does the analysis deal with a paradigm case of causation? Suppose I throw a ball, and the window shatters. My throw (C) causes the window shattering (E) as follows: there is a counterfactual chain of dependence leading from the window shattering (E) to my throw (C), i.e. there are some actual events (the shattering (E), my ball in late flight (D), my ball in early-flight (D1), my throw (C)), such that the shattering (E) counterfactually depends on my ball in late flight (D), and my ball in mid-flight (D) counterfactually depends on my throw (C). Note the counterfactual chains analysis implies that counterfactual dependence between distinct actual events is sufficient for causation (i.e., it implies the truth of SC); for if there is a counterfactual dependence between C and E, then there is a counterfactual chain of dependence between C and E (namely, the chain that only includes C and E). Thus the counterfactual chains analysis squares well with our original intuition that SC really was sufficient for causation.

**2.2.1 Clarifying the notion of 'causal histories'**

Lewis' counterfactual chains analysis renders the causal relation a transitive one. Is causation really transitive? It certainly isn't intransitive. That is, it is clear that there are cases in which C causes D, D causes E, but also C causes E. For instance, the striking of the billiard ball (C) causes the ball's moving across the table (D), and the ball's moving across the table causes the ball to be pocketed (E), but also the striking of the billiard ball (C) causes the ball to be pocketed (E). The choice is then between causation being a transitive relation (such that for all C, D, and E, if C causes D and D causes E, then C causes E), and a non-transitive relation (such that there are cases where C causes D and D causes E in which C is also a cause of E, but also different cases where C isn't a cause of E).  But which? Putative counter-examples to the transitivity of causation are common place. Witness:

An example due to McDermott[40]: Suppose a dog bites off Billy's forefinger (C), which causes Billy to set off a bomb with his left hand (D), which in turn causes the explosion (E). However, surely the dog biting off Billy's forefinger (C) is not a cause of the explosion (E)?

An example due to Hall[41]: Suppose Suzy gives Billy half of a deadly dose of poison (C), which causes Bruno not to give Billy the second half of the deadly dose (D), which in

---

[40] See McDermott [1995]

[41] See Ned Hall, p. 225, [2001]

turn causes Billy's survival (E). However, surely Suzy giving Billy half of a deadly dose (C) is not a cause of Billy's survival (E)?

An example due to Field[42]: Suppose Suzy leaves a bomb outside Billy's door (C), which causes Bruno to deactivate it (D), which in turn causes Billy's survival (E). However, surely Suzy's placing of a bomb (C) is not a cause of Billy's survival (E)?

An example due to Papineau[43]: Suppose Billy's being fat (C) caused him to go on a diet (D), and his going on a diet caused him to be thin (E). However, surely Billy's being fat (C) was not a cause of him becoming thin (E)?

An example due to Bennett[44]: Suppose heavy rain in April (C) caused the delayed electrical storm in June (D), and the delayed electrical storm in June caused the forest fire (E). However, surely the heavy April rain (C) was not a cause of the forest fire (E)?

An example due to Lombard[45]: Suppose Brown's heroic action (C) causes Jones not to be stabbed one night by Smith (D), and that Jones' not being stabbed that night (D) causes the persistent Smith to kill Jones a year later (E). However, surely Brown's heroic action (C) is not a cause of Jones' death a year later (E)?

---

[42] Hartry Field, cited from an unpublished lecture, see Lewis, p. 97, [2001]

[43] See Papineau [1986]

[44] Bennett, p.222-3, [1987]

[45] Lombard, p. 197, [1990]

An example due to Ramachandran[46]: Suppose Singh was due to detonate a bomb but is instead involved in an accident (C) that prevents this. Patel detonates the bomb instead (D) and the bomb explodes (E). However, surely the accident (C) was not a cause of the explosion (E)?

What are we to make of such putative counter-examples to the transitivity of causation? Some have accepted that the causal relation is indeed non-transitive (see Yablo [2004]), some have resisted such counter-examples (see Lewis [1973b]), others have said that there are in fact two different, but equally useful, concepts of causation- one transitive, the other non-transitive - and it depends on context which concept we are using (see Hall [2001]).

Which reaction should we prefer? Lewis, when he defends a transitive causal relation, understands statements of the form 'C causes E' in terms of 'C is in the causal history of E.' He writes: "in rejecting the [putative] counter-examples [to the transitivity of causation]… I think I am doing what Historians do. They trace causal chains, and without more ado, they conclude that what comes at the end of the chain was caused by what went before"[47]. In contrast, the pull against the view that causation is transitive seems to come from the fact that we often understand statements of the form 'C causes E' in terms of 'C causally explains E,' such that when we address the putative counter-examples above, we find it counter-intuitive to cite C as a cause/causal explainer of E; and this is

---

[46] Ramachandran, p390, [2001]

[47] Lewis, p.99, [2001]

supported by the fact that E does not counterfactually depend on C in such examples, confirming the view that counterfactual dependence should hold in paradigm cases of causation.

The matter of whether causation is transitive or not then seems to become an issue of what causal concept we are talking about. We should then understand analyses of causation that render the causal relation transitive to be analysing a causal concept about causal *histories*, and analyses that render the relation non-transitive to be analysing a causal concept about causal *explainers*. This shall subsequently widen the scope of our investigations for an interesting causal concept.

So far we have discussed Hume's counterfactual analysis [1948] and Lewis' counterfactual chains analysis [1973b]. We shall now visit Lewis' later analyses of a more sophisticated variety.

## 2.3 Lewis' quasi-dependence analysis [1986b]

Lewis later thought that what was really required to get round pre-emption cases was not an analysis of causation that required chains of dependence between cause and effect, but merely 'quasi-dependence' between cause and effect. The idea was to drop the requirement that the actual cause or effect be linked by chains of counterfactual dependence in the actual scenario, but merely require that they be linked by a 'sort of/quasi-dependence' by appeal to very similar cases in which there was such a chain of

dependence, which the actual scenario might not possess. The analysis is more formally expressed as follows:

First, we state an analysis of causation that renders the causal relation non-transitive (call it the 'basic quasi-dependence analysis'):

- C causes E iff E quasi-depends on C

Secondly, we state Lewis' quasi-dependence analysis:

- C causes E iff there is chain of quasi-dependence from E to C

Thirdly, we state the truth conditions for chains of quasi-dependence:

- There is a chain of quasi-dependence from E to C iff there are some actual events E, D, D1 etc… C, such that E quasi-depends on D, D on D1 etc… C.

Fourthly, we state the truth conditions for quasi-dependence:

- E quasi-depends on C iff there is an isonomic duplicate of a process from C to E, in which (the counterpart of) E counterfactually depends on (the counterpart of) C.

Fifthly, we state the truth conditions for duplication:

- Two sets of events are duplicates iff they share all the same intrinsic properties.

Sixthly, the truth conditions for isonomic duplication:

- Two sets of events are isonomic duplicates iff they share the same intrinsic properties and laws of nature.

Note that according to the conditions of isonomic duplication, two sets of duplicates or isonomic duplicates can have different extrinsic properties, such that:

- The original set of events and its isonomic duplicate can be the same.
- The original set of events and its isonomic duplicate can be not the same, and exist in the same world.
- The original set of events and its isonomic duplicate can be not the same, and exist in different worlds.

Note also that Lewis does not take *strict* duplication between duplicates for his original quasi-dependence analysis, rather, he allows for duplicates to be duplicates just in case they are intrinsically *similar* to a sufficient degree. So we may express duplication as a coarse-grained notion following Lewis' idea, as follows:

- Two sets of events are duplicates iff they are intrinsically similar

We shall compare the virtues and vices of taking duplication as a strict notion or a coarse notion later. Lewis takes it as a coarse notion, such that to use the analysis to determine whether Suzy's throw caused the shattering in a paradigm case of causation, we might do the following:

> "Take another case, actual or possible, which is intrinsically just like the case of Suzy throwing her rock at the bottle (and which occurs under the same laws)... In this comparison case, we have a causal chain from Suzy's throw to the shattering which does exhibit counterfactual dependence and which is an intrinsic duplicate of the actual chain from Suzy's throw… ([or] near enough)."[48]

It will be worth explaining how the quasi-dependence analysis deals with this case of causation in more detail. Suppose Suzy throws a ball, and the window shatters. Suzy's throw (C) causes the window shattering (E) as follows: there is a chain of quasi-dependence from the window shattering (E) to Suzy's throw (C), that is, there is an isonomic duplicate (in this world, or another similar one) of a process from Suzy's throw (C) to the shattering (E) (namely, an isonomic duplicate of the process which includes all the ball states from and including Suzy's throw up to and including the shattering) in which (the counterpart of) the shattering (E) counterfactually depends on (the counterpart

---

[48] Lewis, p. 82, [2000]. Square brackets are mine.

of) Suzy's throw (C). Thus, the quasi-dependence analysis seems to succeed well in analysing paradigm cases of causation.

Note that the truth of the quasi-dependence analysis relies on the *intrinsicness thesis* being true. The intrinsicness thesis may be expressed as follows:[49]

- For any collection of events S including an effect at t' together with all of its causes from some time t up until t', any isonomic duplicate of S has the same causal structure as S[50]

The idea of the intrinsicness thesis is roughly the idea that 'causal facts are local facts'. Return to our paradigm case, and take the collection of events including the shattering (E), together with all of its causes from and including Suzy's throw (C) up until (E) (i.e. the ball in early (D), mid (D1), and late (D2) flight). The intrinsicness thesis holds that any duplicate of this collection of events with the same laws of nature will have the same causal structure. For this particular case, this seems correct. But we shall visit the vices of the intrinsicness thesis later when we visit cases of prevention and trumping, as it turns out that the intrinsicness thesis is false if we think cases of prevention and trumping are causal. For now, we should state some additional clarifications to the quasi-dependence analysis.

---

[49] The converse is obviously false; if two sets of events share the same causal structure then it is not necessarily the case that they are isonomic duplicates (two sets of events can have the same causal structure yet include different events or laws of nature).

[50] See Hall, p.244 & p. 239, in Collins, Hall and Paul (ed.s) [2001].

## 2.3.1 Clarifying the notion of a 'process'

A preliminary issue with the quasi-dependence analysis regards whether a reductive analysis of process-hood can be given. To take process-hood as an unanalysed primitive (as Lewis seems to do) is unacceptable for three reasons. Firstly, what it is to be a process seems to be a causal notion, and thus if we do not provide a non-causation assuming analysis of process-hood the quasi-dependence analysis seems unacceptably *circular*. Secondly, taking process-hood as primitive would leave the nature of process-hood a persisting mystery, leaving it *unclear* as to when a process exists or not in some complex cases of causation (indeed, this will be a problem for the quasi-dependence analysis when we visit cases of trumping pre-emption). Thirdly, taking process-hood as primitive would defeat Lewis' own project of finding a *reductive* analysis of causation (i.e. reducing causation to relations between non-causal events). Thus, it seems that an analysis of causation must provide an analysis of what it is to be a process, if that analysis is to be *non-circular*, *clear*, and *reductive*.

The quasi-dependence analysis then faces two tough questions which it must answer: Firstly; what is the analysis of process-hood that is required? And secondly, given a successful analysis of process-hood, why not analyse a transitive causal relation in terms of processes simpliciter, i.e. say that C causes E iff there is a causal process running from C to E (or something similar thereabouts), and forget about sophisticated notions like 'isonomic duplication' and 'quasi-dependence'? Lewis is silent on these questions. But

regardless of these issues, it will be fruitful to investigate how the quasi-dependence analysis proposes to treat cases of pre-emption, as such an investigation will provide some helpful morals for a successful analysis of causation which I offer much later.

## 2.3.2 Clarifying 'quasi' and 'would-be' dependence

As we shall see, the strength of the quasi-dependence analysis in dealing with cases of pre-emption will depend on its appeal to cases just like a causal process in the actual case, but without the inclusion of troublesome background events that 'prevent dependence' in a troublesome way, like pre-empted back up alternative causes. Another way of setting out this sort of idea was presented by Collins [2001]. His 'would-be dependence analysis' can be explained as follows:

First, the truth conditions for Collins' analysis:

- C causes E iff there is a chain of counterfactual dependence linking C to E, or there would be such a chain had it not been for some pure dependence preventer P of E on C.

Secondly, the truth conditions for the notion of dependence prevention:

- P is a dependence preventer of E on C iff if P had not occurred then E would counterfactually depend on C.

55

Thirdly, the truth conditions for the notion of *pure* dependence prevention:

- An event P *purely* prevents dependence of E on C iff P is a dependence preventer of E on C, and P prevents no event in a chain of events that would form a counterfactual chain of dependence linking E to C had P not occurred.

Fourthly, a modification of the first analysis gives us Collins' revised analysis:

- C causes E iff there is a chain of counterfactual dependence linking C to E, or there would be such a chain had it not been for some pure dependence preventer in *some not too far-fetched way*.

Collins' analysis is very similar to Lewis' [1973b] counterfactual chains analysis, except it adds the disjunction in the analysis '…or there would be such a chain had it not been for some pure dependence preventer.' The importance of the second disjunct shall become clear when we visit cases of pre-emption. What I shall mention here is the sort of strategy that both the quasi-dependence and would-be dependence analyses employ in order to analyse cases. That is, both employ a strategy of '*getting-rid*' of troublesome background events that prevent dependence of the effect on the cause, in order to obtain counterfactual dependence between an effect and its cause. This strategy shall meet some early successes, but also notable failures when we meet cases of causation that do not satisfy the intrinsicness thesis.

## 2.4 Lewis' Influence analysis [2000]

We now come to Lewis' final analysis of causation; the influence analysis. Lewis' influence analysis can be expressed as follows:

- C causally influences E iff there is a chain of influence from E to C.[51]

First, the truth conditions about chains of influence:

- There is a chain of influence from E to C iff there is a set of actual events E, D, $D_1$ etc… C, such that C influences…, $D_1$ influences D,  etc…. E.

Second, the truth conditions for influence:

- C influences E iff there is a substantial range of $C_1$… $C_n$ of different not-too-distant alterations of C (including the actual alteration of C) and a range of $E_1$… $E_n$ of alterations of E, at least some of which differ, such that if $C_1$ had occurred, $E_1$ would have occurred, if $C_n$ had occurred, $E_n$ would have occurred.[52]

---

[51] Lewis, [2000], p. 190

[52] Lewis, [2000], p. 190

Third, the truth conditions for alterations:

- Let an *alteration* of an event E be either a very fragile version of E or else a very fragile alternative event that is similar to E, but numerically different from E.[53]

To demonstrate how this analysis succeeds in dealing with paradigm cases of causation, consider the following example: Suppose a man poisons himself whilst eating dinner, and that this poison kills its victim more slowly and painfully when taken with food, such that the man would have died a lot more quickly and less painfully if he had just taken the poison without dinner. Following Lewis, we can suppose a collection of alterations (in not-too-distant possible worlds) made to a cause (i.e. the man takes poison (C)) $C_1 \dots C_n$ as follows:

$C_1$ The man takes poison.

$C_2$ The man takes stronger poison.

$C_3$ The man takes poison slightly later.

$C_4$ The man takes poison with painkiller.

$C_5$ The man does not take poison.

We can also suppose a collection of alterations made to the effect in the not-too-distant set of possible worlds (the man dies (E)) $E_1 \dots E_n$ as follows:

---

[53] Lewis, [2000], p. 188

**E$_1$** The man dies.

**E$_2$** The man dies earlier.

**E$_3$** The man dies slightly later.

**E$_4$** The man dies less painfully.

**E$_5$** The man does not die.

Now, to examine whether C causally influences E, we examine whether there is a collection of alterations to C that counterfactually implies a collection of alterations to E (where such alterations occur in the collection of not-too-distant possible worlds). In our list, we can conclude that there is (I have organised the list of C$_n$ and E$_n$ to reflect this). For instance, C$_1$ counterfactually implies E$_1$, C$_2$ counterfactually implies E$_2$, C$_3$ counterfactually implies E$_3$ etc. Thus we have a pattern of counterfactual implication to some alterations of *whether* he dies, *when* he dies, and *how* he dies from some alterations of *whether* he takes the poison, *when* he takes the poison, and *how* he takes the poison. We then conclude that the taking of the poison (C) causally influences the man's death (E), because C influences E; that is, alterations of the man's death are counterfactually implied by alterations of the man's taking poison in the collection of not-too-distant possible worlds.

An important feature of the influence analysis is that it admits of *degrees* of causal influence, and it does so in a very rough and multi-dimensional way. We assess the *degree* of causal influence C has over E in relation to other actual events, such as (A), as

59

follows: We say that if (in the collection of not-too-distant possible worlds) alterations of C counterfactually imply *greater* alterations of E than alterations of A do, then C causally influences E *more* than A does. For example, if (in the collection of not-too-distant possible worlds) alterations of the man taking poison counterfactually imply *greater* alterations to the man's death than alterations of the man eating dinner do, then the man's taking poison causally influences the man's death more than his eating dinner does; and we can see in paradigm cases of this sort that it does. [54]

**2.4.1 Clarifying the notion of 'alterations'**

A first issue with the influence analysis is that Lewis has not given a clear enough definition of what constitutes an 'alteration'; his definition that an alteration of C is either 'a fragile version' or a 'very fragile alternative' of C leaves it unclear as to what these notions actually mean, i.e. it is left unclear what it is to have an altered event which is a version of that same event, and an altered event which is a different event. Choi presents us with some terminological distinctions that help clarify matters:

1. Following Lewis,[55] suppose the *occurrence condition* for an event E is occurrence O in a spatio-temporal region R iff, necessarily, event E occurs iff O occurs in R.

---

[54] Lewis, [2000], p. 188

[55] Lewis, [1986], p. 244-249

2. Following Hempel,[56] suppose when an event E occurs in R it is a *'concrete'* (i.e. fine-grained) event iff its occurrence condition consists of *all* its intrinsic and spatio-temporal properties.

3. Following Strevens,[57] suppose when an event occurs in R it is a *'high level'* (i.e. coarse-grained) event iff its occurrence condition consists of only *some* its intrinsic and spatio-temporal properties.

4. Following Lewis,[58] suppose event C *implies* event E iff, necessarily, C occurs in R iff E occurs in R.

5. Following Choi,[59] suppose a *realiser* of an event C is a concrete event that necessarily implies C.

An example will help illustrate these distinctions: Following (1), the occurrence condition for an event is my saying "Hello" loudly in a spatio-temporal region iff, necessarily, that event occurs iff I say "Hello" loudly in that spatio-temporal region. Following (2), when the event of my saying "Hello" occurs in a spatio-temporal region it is a concrete event iff its occurrence conditions consist of all the intrinsic and spatio-temporal properties that my saying "Hello" consists of (and following (3), it is only a 'high-level' event if its

---

[56] Hempel, [1965], p. 421-423

[57] Strevens, [2003], p. 398-399

[58] Lewis, [1986], p.255

[59] Choi, [2005], p.102

occurrence conditions only consist of some of those properties). Following (4), the event of my saying "Hello" loudly implies the event of my saying "Hello" simpliciter because the event in which the former occurrence occurs necessitates the event in which the latter occurrence occurs. Following (5), when I say "Hello" loudly this is a concrete event that implies my saying "Hello" simpliciter, because given the occurrence condition for the former event consists of all the intrinsic and spatio-temporal properties of R, the latter event occurs in all regions in which the former occurs; my saying "Hello" loudly is thus a realiser for my saying "Hello" simpliciter.

Choi observes two things: Firstly, that one concrete event can realise more than one high-level event. So for example, the concrete event of my saying "Hello" loudly and clearly implies and thus realises the event whose occurrence conditions are my saying "Hello" simpliciter, the event whose occurrence conditions are my saying "Hello" loudly, and the event whose occurrence conditions are my saying "Hello" clearly. Secondly, we observe that one high level event can have more than one concrete realiser. For example, a set of possible concrete events (as well as the actual concrete event) of my saying "Hello" implies and thus realises the high level event whose occurrence condition is my saying "Hello" loudly or clearly.

The third set of truth conditions for Lewis' influence analysis (above), can then be more accurately expressed as follows:

- Let an *alteration* of event E be either a very fragile version of E (*and so a realiser*

*of E*) or else a fragile alternative event that is similar to E but numerically different from E (*and so a fragile event that is not a realiser of E but similar to realisers of E*).

To return to our example, an alteration to the event of my saying "Hello" loudly can be either a very fragile version of my saying "Hello" loudly, such as my saying "Hello" loudly but in a different way, or else a fragile alternative event similar to my saying "Hello" loudly, but numerically different from my saying "Hello" loudly, such as my saying "Hello" slightly less loudly.

### 2.4.2 Clarifying the notion of 'not-too-distant'

A second issue with Lewis' influence analysis concerns what it is for possible alterations made to the cause to be 'not-too-distant.' Suppose Billy and Suzy both throw rocks at a window, and Suzy's throw causes the shattering of the window just before Billy's would have. Now, altering Billy's throw such that it occurs a little earlier makes no difference to the shattering (or a negligible difference), but altering his throw so that it occurred a great deal earlier means that Billy's rock caused the smashing by beating Suzy's to the window. We want the analysis to tell us that Suzy's throw was the cause of the smashing. So presumably, the set of 'not-too-distant' worlds must be wide enough to include the alterations made to the effect made by alterations of the cause (such as Suzy's throw), but narrow enough to exclude the alterations made to the effect by alterations of non-causes (such as Billy's throw). Thus a clear demarcation between what constitutes "not-too

distant" as opposed to "too distant" is essential to providing a clear demarcation between what constitutes non causal influencers as opposed to causal influencers.

The problem, as McDermott observes, is that "too distant means… what? It begins to look as if it just means distant enough to give the wrong answer."[60] The criticism here is that there seems no principled way to demarcate which worlds are not-too-distant from which worlds are not. Instead it seems we take the set of worlds which gives us the intuitively correct results we want in the first place and so we beg the question of how we can explain C as a cause or causal influencer of E in a non circular way.

The solution: Presumably the close set of worlds in which alterations of Billy's throw *do not* alter the shattering (or do so negligibly), is the set of not-too-distant worlds by which we want to determine Billy's throw as having no or little causal influence on the shattering in the actual world. This set of worlds is identified because it stands in firm contrast to the relatively closer set of worlds in which alterations to Suzy's throw *do* alter the shattering a great deal; there is, as Maslan puts it, a 'relatively inert zone',[61] i.e. a 'zone' of worlds in which Billy's throw is causally inert with relation to the shattering. Thus, we can now determine whether a set of possible worlds is too distant or not, by asking whether those worlds contain closer worlds which are 'relatively inert' with respect to altering E. If there is such a distinguishable class of relatively inert worlds, then we are in a position to say that the class of worlds is too distant, and if not, we are

---

[60] McDermott, [2002], p.91

[61] Maslan, [2004b]

not.

## 2.4.3 Clarifying the notion of 'comparative alteration'

As Maslan observes,[62] a problem with comparing the degree of alteration (between two altered events or two alterations to one event), is that some alterations may not even be comparable; some cases may just be impossible to judge. For instance, with respect to the earlier example, there seems no way to assess whether E2 (the man dying earlier) is more or less of an alteration of E1 (the man dying) than E4 (the man dying less painfully). Lewis does not anticipate this problem.

The answer often seems to lie in the particular context of the inquiry. As Maslan puts it; "the process of selecting the decisive factor among many [as to what alteration to an event alters that event more than another alteration] depends heavily on the interests of the speaker."[63] But this said, there is nothing to guarantee that there will be an appropriate context that will determine the decisive factor for us in all cases. For example: It is hard to see what context will determine the decisive factor in what most alters a yellow square; when it is altered into a red triangle, or altered into a blue pentagon? Each colour is an equal distance away from each other on the colour scheme, and either alteration either adds or removes one single side to the original shape. It then seems that there will be undecidable cases.

---

[62] Maslan, [2004b], p. 582-4

[63] Maslan, [2004b], p. 582-4, square brackets are mine.

Examples of this form generate problems for the influence analysis insofar as there seem to be cases in which it seems impossible to (usefully) compare a distance of (not-too-distant) alterations of one event with an *equal* distance of (not-too-distant) alterations of another, and thereby cases in which it is impossible to assess whether one cause causally influences the effect more than another. The influence analysis then suffers from extreme vagueness; it can often only give us very vague approximations as to whether one event causally influences more than another or not. A further example will illustrate this point: Suppose we both shatter the same window simultaneously, I use a magic spell to do so, and you throw a rock at it: which is the greater causal influencer of the shattering? Presumably we want to say they are both equal causal influencers of the shattering; however, the influence analysis provides us with no machinery by which to determine this, that is, there seems no palpable metric by which we can compare the distance between alterations made to a magic spell and alterations made to a throw. This problem for the influence analysis remains unresolved.

**2.4.4 Clarifying the notion of 'causal influence'**

It has been criticised that the 'influence' analysis is no longer one of causation, but of an altogether new concept. For instance, Collins writes:

> "This new theory of causal influence amounts to a change in topic. Lewis is not
>
> offering a new answer to the old question: "What is it for this event to be a cause

66

of an event?" Rather, he is offering an answer to a quite new question: "What is it for this event to have a causal influence on that event? … As it stands, the new theory lacks the resources to identify an event's causes from among those things which merely had some causal influence on it."[64]

Is it really a change in topic? As Hitchcock observes,[65] propositions of the form 'C influences E' may be better synonymous with 'C makes a difference to E' or 'C has an effect on E' than 'C causes E.' But as Maslan counters, "treating influence as an interesting causal concept in its own right justifies overlooking this objection."[66] Indeed, "there seems to be a number of different causal concepts that we do not always clearly distinguish from each other,"[67] and influence seems to be tied up with some of them. Given this, it seems as if the influence analysis is not a clear departure from our platitudes about causation to warrant dismissal from the overall project of finding an interesting counterfactual analysis of causation.

I have mentioned two issues with the influence analysis so far; firstly that it fails to determine causal influence when it comes to comparing the causal efficacy of dissimilar events which alter an effect in a dissimilar way, and secondly that it might amount to a change in topic. Otherwise I have discussed one issue about what determines the set of not-too-distant possible worlds, and one about what determines alterations. These

---

[64] Collins, [2000], p. 231

[65] Hitchcock, [2002], p. 384

[66] Maslan, [2004b], p. 593

[67] Maslan, [2004b], p. 593

clarifications will play a significant role in how effectively the influence analysis copes

with the problematic cases to come.

**Chapter 3: The Deterministic Pre-emption Cases**

**3.0 What is pre-emption?**

So far I have introduced the analyses of causation I wish to compare, and some basic problems that confront them. We now come to an altogether different problem - the problem of pre-emption. What are cases of pre-emption? Cases of pre-emption present the most harassing problems for counterfactual analyses of causation. In short, cases of pre-emption will show how counterfactual dependence, and many counterfactual analyses of causation, are not *necessary* for causation. I offer some different and general characterisations of pre-empting and pre-empted causes here.[68]

- C is a pre-empting cause of E iff C causes E, but had C not caused E, something else would have
- C is a pre-empted cause of E iff C does *not* cause E, but would have had something else not caused E

Cases of pre-emption are cases of redundant causation. Lewis defined redundant causation in PostScript E of 'Causation':

"Suppose we have two events c1 and c2, and another event e distinct from both of them; and in actuality all three occur; and if either one of c1 and c2 had occurred

---

[68] Maslan, [2004b], p. 582

without the other, then also e would have occurred, but if neither c1 nor c2 had occurred, then e would not have occurred. Then I shall say that c1 and c2 are redundant causes of e." (Lewis, [1986], p. 193)

Lewis' definition here may be put more formally[69], as follows: Where Q, E and C are distinct actual events, then;

Q is a redundant cause of E iff

- $(o\text{Q} \,\&\, {\sim}o\text{C}) \,\square\!\!\rightarrow o\text{E}$

- $({\sim}o\text{Q} \,\&\, o\text{C}) \,\square\!\!\rightarrow o\text{E}$

- $({\sim}o\text{Q} \,\&\, {\sim}o\text{C}) \,\square\!\!\rightarrow {\sim}o\text{E}$[70]

Maslan also defined redundant causation in terms of 'redundancy sets,' as follows:[71] A redundancy set for an effect E may be defined as a set of distinct events {C1… Cn} such that:

- If no member of the set {C1… Cn} had occurred, then E would not have occurred.

---

[69] As is standard, '$\square\!\!\rightarrow$' is counterfactual implication, '$o$' is a predicate of occurrence, '~' and '&' are classical negation and conjunction respectively.

[70] 1. If Q had occurred and C hadn't, then E would have occurred. 2. If Q hadn't occurred and C had, then E would have occurred. 3. If neither C nor Q had occurred, then E wouldn't have occurred.

[71] Maslan, [2004].

- It is not the case that if only one member of the set had not occurred, then E would not have occurred.

- If one member of the set had occurred without the other members, then E would still have occurred.

- Pre-empting causes are members of the set that are causes of E. Pre-empted alternative causes are members of the set that are not causes of E.

- If more than one member of the set is a cause of E, then they are over-determiners (or double pre-empters).

Cases of deterministic pre-emption come in different general types, the ones to be found in the literature I taxonomise as follows:

1. Early pre-emption

2. Late pre-emption

3. Middle pre-emption

4. Double pre-emption

5. Bunzl pre-emption

6. Trumping pre-emption

7. Pre-emptive prevention

8. Double prevention

71

We now come to a major part of this thesis, that of investigating the cases of pre-emption[72] and testing the prominent analyses we explored in chapter 1 against them. Some of the cases we shall meet will be quite complex. We should proceed clearly and methodically; following this ethos, I shall proceed in each section to give the following:

i)      An explanation of the case considered

ii)     A diagrammatic re-construction of the form of the case considered[73]

iii)    Examples of the form of the case considered

iv)     Demonstrations of how analyses succeed/fail in analysing the case considered

v)      A scorecard cataloguing the successes of the analyses considered.

We are now ready to proceed.

## 3.1 Early pre-emption

---

[72] A general characterisation of which was given at the beginning of this thesis.

[73] Explanation about the diagrams is given here: Grey outlined circles represent actual events, white outlined circles represent actual omissions; their dashed counterparts represent non-actual events/omissions. Black arrows represent actual causal connections, black blobbed lines represent actual causal connections of prevention; their dashed counterparts represent non-actual causal connections. A long black splitting line represents a fact which inhibits an event from causing E; its dashed counterpart represents a fact which would (but actually doesn't) inhibit an event from causing E. A big black arrow represents a 'trumping' causal connection; its dashed counterpart represents a non-actual 'trumping' causal connection.

*Intuitive explanation*: In cases of early pre-emption, the causal process running from the pre-empted 'back-up' cause is inhibited from causing the effect *before* the main process running from the pre-empting cause has gone to completion.

*Diagrammatic reconstruction*: Consider figs. below. The first represents what occurs in the actual world in which C occurs: C causes E via D. Additionally, A would have caused B, and B would have caused E, had C not occurred to inhibit B from causing E (such that C pre-empts the process running from A to E in the actual world, see the second diagram).
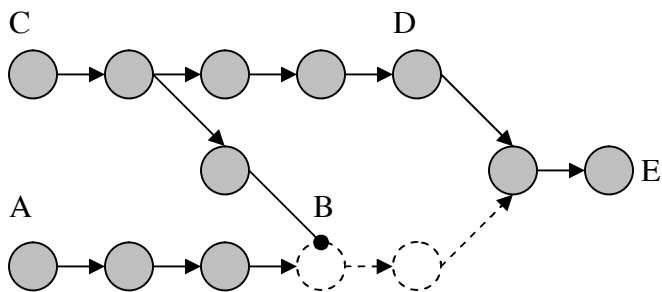


The actual world: C causes E via D, C prevents B from there being a causal chain from A to E
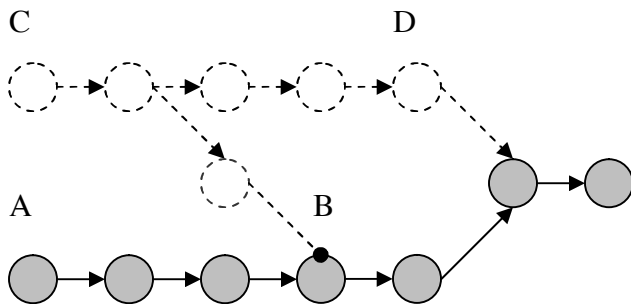
The closest possible worlds in which C does not occur: B, no longer prevented, occurs, and there is a causal chain from A to E

### 3.1.1 Coarse-grained early pre-emption

**Example (1):** An example in which the coarse-grained version of the effect still would have occurred had the cause not occurred. Call this type of case 'coarse-grained early pre- emption.'

SUZY AND BILLY#1 (an example from Lewis[74]): Suppose Suzy and Billy are playing with rocks near a window. Suppose Suzy throws a rock at the window (C), sending it through the air in mid trajectory (D), resulting in the window's shattering (E). Suppose also that Billy had the intention to throw his rock at the window (A) (for ease, call this event Billy's 'near throw'), which would have in turn caused him to throw his rock (B), resulting in a similar (albeit slightly different) shattering (E), had it not been for the fact that Suzy's throw (C) prevented Billy from making his throw (B) (here we suppose that Billy felt no compulsion to throw his rock given he saw Suzy throw hers). We conclude that Suzy's throw (C) caused the shattering (E), and that Billy's near-throw (A) was the *pre-empted* back-up cause.

---

[74] See Lewis, [1973]

**Example (2)**: Another example of 'coarse-grained early pre-emption.'

CAR RACE#1 (an example from Kvart[75]): Suppose a sports car races against a wagon from a starting line to a finishing ribbon. Suppose the wagon starts off on its route (C), passes through the half way mark (D), and breaks the finishing ribbon (E), winning the race. Suppose also that the sports car starts off on its route (A), and would have driven past the half way mark (B) had it not been for the fact that the wagon starting off on its route (C) prevented the sports car from continuing the race after the half way mark (B) (here we suppose the wagon starting off on its route activated a barrier that served to block the car from continuing the race at some early juncture). We conclude that the wagon starting off on its route (C), caused the breaking of the finishing ribbon (E), and that the sports car starting off on its route (A) was the pre-empted back up cause.

How do our analyses deal with these examples?

*The basic Humean analysis fails*: Recall the details of the basic Humean analysis:

- C causes E iff E counterfactually depends on C, C precedes E, and C and E are distinct actual events.

The basic Humean analysis fails to successfully analyse the first example. As follows:

---

[75] Kvart, [2000], p. 411

1. In terms of counterfactuals: The coarse-grained version of the shattering of the window (E), does not counterfactually depend on Suzy's throw (C). Or in other words: If Suzy's throw (C) had not occurred, a coarse-grained version of the shattering of the window (E), still would have occurred (albeit in a slightly different time and manner), because Billy's rock would have sufficed for a similar shattering.

2. In terms of Lewisian counterfactual semantics: The most similar possible worlds in which Suzy did not throw her rock (C), are all worlds in which coarse-grained versions of the shattering (E) still occur, because Billy's throw suffices for a similar shattering in such worlds.

The basic Humean analysis fails to successfully analyse the second example. As follows:

1. In terms of counterfactuals: The coarse-grained version of the breaking of the finishing ribbon (E), does not counterfactually depend on the wagon starting off on its route (C). Or in other words; if the wagon starting off on its route (C) had not occurred, a coarse-grained version of the breaking of the finishing ribbon (E), still would have occurred (albeit in a slightly different time and manner), because the fast car would have sufficed for a similar breaking of the finishing ribbon.

2. In terms of Lewisian counterfactual semantics: The most similar possible worlds in which the wagon did not start off on its route (C), are all worlds in which coarse-grained versions of the breaking of the finishing ribbon (E) still occur,

because the fast car would have sufficed for a similar breaking of the finishing ribbon in such worlds.

Is there a way to save something like a version of the Humean analysis? A quick fix is offered here[76]:

*The fine-grained Humean analysis mk. 1 succeeds:* A quick fix to the Humean analysis would be to take advantage of the notion of *fragile/ fine-grained* events we investigated in (1.2). Recall,

- An event is fragile/fine-grained iff all its properties are essential to it, and thus could not have occurred in any other time or manner.

Taking this notion of fragility into account, we might generate a revised version of the basic Humean analysis. We might state this revised theory as follows (call it 'the fine-grained Humean analysis mk.1'):

- C causes E iff E counterfactually depends on C, and C and E are actual distinct events (where E is a fine-grained event).

The fine-grained Humean analysis mk. 1 successfully analyses the first example, as follows:

---

[76] This idea is briefly considered and dismissed by Lewis in his [2000a].

1. In terms of counterfactuals: The fine-grained version of the shattering of the window (E) counterfactually depends on Suzy's throw (C), or in other words: If Suzy's throw (C) had not occurred, the fine-grained version of the shattering (E) (consisting of the precise time and manner in which the shattering actually occurred), no longer would have occurred, given Billy's throw (A) would have sufficed for a different, albeit similar shattering (i.e. an event which is not (E) but similar to (E)).

2. In terms of Lewisian counterfactual semantics: The most similar possible worlds in which Suzy does not throw her rock (C), are all worlds in which the fragile version of the shattering (E) does not occur, because Billy's throw (A) suffices for a different albeit similar shattering in such worlds (we may suppose it would have occurred slightly later, and less rigorously because of Billy's inferior strength).

This fine-grained Humean analysis mk. 1 successfully analyses the second example, as follows:

1. In terms of counterfactuals: The fine-grained version of the breaking of the finishing ribbon (E) counterfactually depends on the wagon taking its route (C). Or in other words, if the wagon starting off on its route (C) had not occurred, the fine-grained version of the breaking of the finishing ribbon (E) (consisting of the precise time and manner in which the ribbon-breaking actually occurred), no

longer would have, given the race car (A) would have sufficed for a different, albeit similar breaking of the finishing ribbon.

2. In terms of Lewisian counterfactual semantics: The most similar possible worlds in which the wagon does not start off on its route (C), are all worlds in which the fragile version of the ribbon breaking (E) does not occur, because the race car would have sufficed for a different albeit similar ribbon-breaking in such worlds (we suppose it would have occurred earlier, given the car would not have been blocked by the barrier and would have accelerated to very fast velocities on the finishing stretch, and would have occurred more rigorously because of the race car's finishing speed).

The fine-grained Humean analysis thus successfully handles our two basic examples of early pre-emption. However, there are two main problems with the revised analysis:

*First problem*: It is a contingent matter of fact that close physical bodies in our world exert some gravitational pull on other close physical bodies (no matter how minute). Let us consider the first example in this light. It then turns out that the presence of Billy was a cause of the shattering as well according to our analysis. As follows; had Billy *not* been present at Suzy's throw, then the trajectory of Suzy's throw would have been *minutely* altered due to the lack of minute gravitational forces on that trajectory that Billy's presence was responsible for. According to the fine-grained Humean analysis, we end up with the conclusion that Billy's presence was also a cause of the shattering (which we may presume is a questionable conclusion!). It is also clear that, upon similar reasoning,

the presence of other intuitively innocuous events would have ended up being labeled as causes (such as the presence of other pebbles on the ground, the precise rotation of the planet Mars around the sun, and the presence of the roof on the building that houses the window etc.).

Are these causal claims spurious ones? No. There is causation between coarse-grained events, causation between fine-grained events, and permutations between the two. For instance, there is causation between the coarse-grained events of the throw and the shattering, and also causation between fine-grained events of the throw and the shattering, and permutations between the two. Whether a causal proposition such as 'the throw caused the shattering' is a proposition about coarse-grained causation or fine-grained causation is dependent on what terms like 'the throw' and 'the shattering' are understood to denote; expressions such as 'Suzy's throw' and 'the shattering' can then be ambiguous as to what sort of entity they express (that is, ambiguous between coarse-grained and fine-grained entities). Accordingly, the claim that Billy's presence was a cause of the shattering is true, so long as we understand the relata as fine-grained entities, but false, if we are considering them as their typically understood coarse-grained entities; and a disambiguation procedure will clear up exactly whether claims of the form 'Billy's presence caused the shattering' are true or not. To generalise; whether a causal claim is true or not can then be clarified by what the causal relata are taken to be in each case. Reasons for thinking that the claims we considered above are *always* spurious are reasons that confuse the claim as being *always* a claim about causation between coarse-grained events, when the claim is not spurious as a claim about causation between fine-grained

events. The mistake is understandable - in ordinary contexts we usually talk about causation between coarse-grained events and not fine. But I say that fine-grained causation is still a valid category permissive of sound causal claims. Furthermore (and to strengthen this view), the above causal claims, that otherwise might seem intuitively spurious, can be given a context in which they sound wholly intuitive and even appropriate. As follows:

Clearly, the claim that Billy caused the shattering *would* be spurious from the more ordinary context of the owner of the window who wanted to know the major causal contributor of the *coarse*-grained event of the shattering. But, on the other hand, clearly the claim that Billy's presence caused the shattering would *not* be spurious in the context of the fastidious micro-physicist interested in analysing the gravitational effects, no matter how weak, that physical bodies exert on other physical bodies. It then seems clear that what we say about what causes coarse-grained versions of events is often going to be very different from (and in less abundance than) what we say about what causes fine-grained versions of events, and that both types of causal claims are appropriate to some context or other, and furthermore appropriate to feature as true causal claims. The criticism that the above analysis generates false causal claims is then misguided.

So an analysis of causation that appeals to fine-grained events as its relata, upon deeper inspection, does not obviously generate false causal claims. Where it does fail is that it tells us that there is *no* coarse-grained causation at all, given its explicit restriction of

causation to fine-grained events. As a consequence, the analysis is *incomplete* if it is to be taken as a *generic* analysis of causation. This is then a first failing of the analysis.

*Second problem*: The second reason why our revised Humean analysis fails is that it doesn't even do the job it was designed to do. Namely, it can't successfully analyse all cases of early pre-emption. To see this, we need but consider cases in which the back-up pre-empted cause would have caused the exact same fine-grained version of the effect which the actual pre-empting cause causes! It will be illuminating to explain how this could be the case by the following (perhaps contrived, but nonetheless logically possible) example:

### 3.1.2 Fine-grained early pre-emption

*The fine-grained Humean analysis Mk. 1 fails upon a revised example.*

**Example (3):** An example in which the fine-grained version of the effect still would have occurred had the cause not occurred. Term this type of case 'fine-grained early pre-emption'.

BILLY AND SUZY#2:[77] Suppose again, that Suzy and Billy are playing with rocks near a window, and that Suzy throws a rock (C), causing the shattering of the window (E), but this time, suppose that if Suzy had not thrown her rock, then Billy would have thrown his

---

[77] This example is a tweaked version of an example suggested by Yablo. See Collins, Hall and Paul [2001].

'smartrock' (A) half a second later. Suppose Billy's smartrock is a highly sophisticated piece of equipment, which is programmed to cause the exact same shattering (at the precise same time and manner) as Suzy's rock had he thrown it half a second after Suzy's actual throw (suppose it would do this by accelerating at very high speeds towards the window, then de-accelerating to a velocity identical to Suzy's rock in its actual flight, and then entering the window in the exact same manner and time as Suzy's actual rock does). Thus, had Suzy not thrown her rock (C), half a second later Billy would have thrown his smartrock (A), and the fragile version of the shattering (E) still would have occurred.

**Example (4):** Another example of 'fine-grained early pre-emption'.

CAR RACE#2:[78] Suppose a similar case to that of CAR RACE#1 (example 2), except for the following details: Suppose two race cars, intrinsically identical except for their colour, are in a race to the finishing ribbon, and that the red car starting (C) causes the ribbon breaking (E), and that it was the red car that activated a barrier that blocked the blue car from continuing on its route (B). Suppose also that if the red car starting (C) had not occurred, then the blue car starting (A) would have caused an exact same ribbon-breaking (so suppose this would have happened because the blue car would not have been blocked by a barrier, and would have accelerated in the later part of the race to, by complete luck, cause an exact same ribbon-breaking). Thus, had the red car not started on its route (C), the starting of the blue car (A) would have caused the fragile version of the ribbon-breaking (E).

---

[78] See Kvart, [2000], p. 411-2, for a similar example.

It is easy to see how the fine-grained Humean analysis fails in each example. In the first example, the fine-grained version of the shattering does not counterfactually depend on Suzy's throw, which is thence not analysed as a cause of the shattering. In the second example, the fine-grained version of the ribbon breaking does not counterfactually depend on the red car starting, which is thence not analysed as a cause of the ribbon breaking. The fine-grained Humean analysis thus fails to handle such fine-grained early pre-emption examples.

Should we discount the above two examples? Perhaps it might be complained that such examples are too contrived or are impossible given our actual laws of nature, and are thereby not serious counter-examples to an analysis of causation. Indeed, perhaps it turns out, when we have finalised a theory about all the laws of nature, that in our world no such cases are nomologically possible. This would be a worthy complaint if this were true, and all we were interested in was analysing a contingent notion of fine-grained or micro-physical causation between events; but the analysis would still be no good in analysing the concept of fine-grained causation as it is in *all logically possible worlds*, and the project of analysing the notion of causation in all logically possible worlds is a project we would ideally like to succeed in. This not only raises the issue of what we want from an analysis of causation, but also which examples we are permitted to use to test it against. Much can be said about this subject, but we shall proceed as if it is the concept of causation that we wish to analyse, and the examples we are permitted to use to test it are those ones that are  logically possible, well explained, intuitive, and clear.

So far, and with respect to fine tuning Hume's counterfactual analysis, we have considered 'fine graining the effect' in order to deal with the pre-emption cases. What about a different, albeit similar strategy, of fine graining the cause? Following this thought we might say:

- C causes E iff E counterfactually depends on C, and C and E are actual distinct events (where C is a fine-grained event, and E is a coarse-grained event).

Call this analysis 'the fine-grained Humean analysis mk.2'. This analysis doesn't even handle basic cases of causation (which do not involve pre-emption) correctly. Suzy throws a rock (C), it shatters the window (E), if the fine-grained version of Suzy's throw (C) had not occurred, a very similar (albeit different) version of her throw (C') would have, and thus the coarse-grained version of the shattering (E) still would have occurred. Suzy's throw is therefore not analysed as a cause of the shattering; the incorrect analysis.

Yet a third option might be to 'fine grain the effect and cause' together with respect to Hume's counterfactual analysis. That is, we might say:

- C causes E iff E counterfactually depends on C, and C and E are actual distinct events (where C and E are fine-grained events).

Call this analysis 'the fine-grained Humean analysis mk. 3'. Does this analysis help us? It is clear that this analysis suffers from the first main problem we considered with its mk.1 counterpart - that is, the analysis fails to accommodate causation about coarse-grained entities and thus provides us with an incomplete analysis. Can it deal with early pre-emption cases any better? It successfully deals with all the early pre-emption cases so far (a brief re-examination of the examples easily demonstrates this). But, unfortunately, there are basic cases of causation (which do not involve pre-emption) which the analysis fails to handle correctly. Consider the following example:

**Example (5)**: FAIRIES (my own example). Suppose Suzy's throw of a rock (C) causes the shattering of the window (E). Suppose it does this in quite a fantastical way; Suzy throws her rock, and a team of fairies intercept the ball, and guide it along a particular trajectory, smashing it through the window. Suppose that the fairies were intent on the shattering of the window occurring in a very particular time and manner, such that if Suzy had thrown earlier or softer, or later or harder, then the fairies would alter their intermediary flight accordingly, such they would fly some periods more quickly or less quickly according to what the best way to cause the same fine-grained version of the shattering would be.

As contrived as this example may be, it shows up an important fact; the fine-grained Humean analysis mk. 3 fails to analyse this case of causation correctly. As follows: If the fine-grained version of Suzy's throw (C) had not occurred, then her throw would have occurred slightly later, earlier, a bit harder, or softer, and also the fairies would have

altered their guidance of the ball in its intermediary flight appropriately such that the fine-grained version of the shattering (E) still would have occurred. Accordingly, Suzy's throw is not analysed as a cause of the shattering in the example; the incorrect analysis.

It is clear from this section that the 'fine graining' strategy won't help us in dealing with pre-emption cases. We now move onto a different strategy suggested by Lewis' first analysis of causation:

*Lewis' (1973) counterfactual chains analysis succeeds in the examples so far.* Recall some of the details of Lewis' 1973 analysis, where C and E are distinct actual events:

- C causes E iff there is a counterfactual chain of dependence leading from E to C.
- There is a counterfactual chain of dependence from E to C iff there is a set of actual events E, D, D1 etc… C, such that E counterfactually depends on D, D on D1 etc…. C.

This analysis succeeds in analysing our previous examples correctly. For simplicity's sake, we consider just either of examples (1) and (3). As follows:

- There is a counterfactual chain of dependence leading from the shattering (E) to Suzy's throw (C), or in other words, there is a set of actual events consisting of the shattering (E), the ball in mid-flight (D), and Suzy's throw (C), such that the

shattering (E) counterfactually depends on the ball in mid-flight (D), and the ball in mid-flight (D), counterfactually depends on Suzy's throw (C).

Thus, Suzy's throw is analysed as cause of the shattering in both examples; the correct analysis. The success of the analysis relies on the strategy of appealing to 'chains' of dependence; where there is no direct counterfactual dependence between the cause and effect, the cause and effect are linked by a chain of counterfactual dependence.

We have already seen why the shattering does not counterfactually depend on Suzy's throw (C). But why does it counterfactually depend on the ball in mid-flight (D) as we just assumed? The answer comes from the ruling out certain backtrackers; namely, from ruling out the backtracker 'if the ball's being in mid flight had not occurred, Suzy's throw wouldn't have'. Rather, if the ball's being in mid flight (D) had not occurred, Suzy's throw (C) would still have occurred - and thus Billy's non-throw would have still occurred (recall, Billy throws only if Suzy doesn't in our example), and thus the shattering (E) wouldn't have occurred. How then are such backtrackers ruled out? We noted earlier that the Lewisian strategy for ruling our backtrackers was hardly complete, and that some other strategy would have to be employed in order to give us an analysis of the requisite forward tracking counterfactuals here. We shall not spend any time on this, for even on the supposition that the Lewisian strategy for ruling out backtrackers works, Lewis' analysis still has many other problems.

As we saw earlier, Lewis' appeal to chains of dependence renders the causal relation transitive. An analysis of a transitive causal relation, as we also noted earlier, is fine so long as we want to have an analysis of causation that tells us when an event is linked in the causal history of a certain effect, but not fine if we want the analysis also to tell when an event provides a causal explanation for an effect (as we saw earlier, not all causal histories give causal explanations). This point aside, we turn our attention to whether the counterfactual chains analysis can succeed (in the relevant sense of providing us with an analysis of 'causal histories') in all cases of early pre-emption.

As it turns out, it can't, on the premise that there can be causal action at a 'spatial distance'. Action at a spatial distance is a proposed type of causation that involves an event causally acting on another event but over a spatial distance, and not through any spatially intermediary events. To see how a scheme for an early pre-emption counter-example to Lewis' counterfactual chain analysis can be generated by appeal to this notion[79], consider the form of our first early pre-emption example (1), except for the following: Firstly, we remove causal intermediaries from C to E (including event D), secondly we maintain that C causes E directly, and thirdly we deny a principle about locality (which holds that causes are connected to their effects via spatio-temporally contiguous sequences of causal intermediaries) and assume that C causes E at a spatial distance.

---

[79] See Collins, Hall, and Paul, [2001], p. 24

We can already see how this sort of example will prove troublesome for Lewis' [1973] analysis, since the success of his treatment of early pre-emption cases relies on there being causal intermediaries running from cause to effect in order for there to be a counterfactual chain from the cause to the effect. So it seems that if we can find an example in which no such intermediaries occur then Lewis' analysis can't hope to succeed. Here is an example that illustrates the form of this proposed counter-example we have just considered:

### 3.1.3 Non-local early pre-emption

**Example (6):** An example in which the effect still would have occurred had the cause not occurred, but where C causes E at a spatial distance. Term this type of case a case of 'non-local pre-emption'.

SUZY AND BILLY#3 (a variation of an example due to Maslan[80]): Consider example SUZY AND BILLY#1 again (Example 1) of early pre-emption, except for the following details: Firstly, Suzy is a witch with magical powers, and is responsible for the smashing of the window not by throwing a rock at it, but by causing it directly with a magic spell; secondly, we suppose the magic spell smashes the window at a spatial distance without appeal to any causal intermediaries. We maintain the premise that had Suzy not caused the shattering Billy's throw would have caused it.

---

[80] Maslan, 2004, p. 598

*Lewis' counterfactual-chains analysis fails* in this revised example, because there is no chain of counterfactual dependence linking the smashing and Suzy's spell. As a consequence, if the spell (C) had not occurred, then Billy's throw (B) would have caused the smashing (E). Thus Suzy's spell (C) is not analysed as a cause of the smashing; the incorrect analysis.

How respond to this example in defence of Lewis' counterfactual chains analysis? Firstly, we might be tempted to assert that spatial action at a distance is (at least nomologically) impossible, and thereby hold that action at a distance is *good for nothing* (and that we consequently shouldn't care about trying to analyse whatever alleged causation there is going on in such cases). Secondly, we might restrict what Lewis' analysis *is good for*. Namely, we might just say that Lewis' analysis analyses causation in worlds without action at a spatial distance (like worlds which have what we take as our laws of nature for instance). We shall have to leave these options unexplored, as our main target is to analyse causation as it occurs across all logically possible worlds, with respect to all logically possible and intuitively clear cases; and Lewis' analysis fails to handle our example on these grounds.

We shall now move on to a novel, non-Lewisian sort of revision to the basic Humean analysis.

*The holding-fixed analysis succeeds* (a proto-type analysis of my own): What is problematic about our above examples of early pre-emption is the fact that had the cause

not occurred, it would have given way to a back-up causal process that would have suffced for either a coarse or a fine-grained version of the effect in each case. An obvious response to this problem would be to simply find some way in which we *don't let the causal process give way* to the pre-empted causal process had the cause not occurred. This is a strategy suggested by Yablo.[81] We shall differ in the details, but we can set out our basic idea by revising the basic Humean analysis as follows; where C, E, and F are distinct actual events, we might say:

- C causes E iff there is some event F such that if F had still occurred and C not, then E would not have occurred.

Call this the '*holding-fixed* analysis',[82] (my own analysis) simply because the strategy the analysis employs is to 'hold fixed' some of the troublesome background pre-empted causes that seem to create all the trouble, and only then test for counterfactual dependence between cause and effect. We should first give a Lewisian style semantics for our new counterfactual notion. As follows:

- There is some event F such that if F had still occurred and C not, then E would not have occurred is true iff there is some event F (or some sum of events F) such

---

[81] See Yablo [2001].

[82] For the sake of simplicity, I will be ignoring cases in which there are no actual events distinct from C and E, but in which the former causes the latter. The interest in the 'holding-fixed' analysis is purely in terms of what morals we may draw from the strategy it employs in dealing with cases.

that the closest possible F occurring and not-C occurring worlds are not-E occurring worlds.

How does the holding-fixed analysis fare up against the above Billy and Suzy examples? We can deal with all of them swiftly in one go:

The analysis tells us Suzy was a cause of the shattering in each case, as follows:

- In terms of counterfactuals: There is some event (F) (Billy holding onto his rock throughout) such that if F had still occurred, and Suzy's throw/spell (C) had not occurred, then the smashing (E) would not have occurred.
- In terms of Lewisian-style semantics: there is some event (F) (Billy holding onto his rock throughout) such that the closest possible worlds in which F occurs and Suzy's throw/spell (C) does not are worlds in which the shattering (E) does not occur.

The success of the analysis relies on the fact that we can 'hold fixed' certain events - in this case, it is the event in which Billy holds onto his smartrock throughout Suzy's throw and the shattering, and it is this 'holding fixed' that prevents the cause from giving way to the pre-empted back-up cause. A first question we might ask of this strategy is: why hold an event (as we set them up in section 3) fixed and not, say, facts (true propositions about states of affairs)? Or, in other words, why not have an alternative analysis that looks like the following:

- C causes E iff there is some fact F such that if F had still obtained and C not occurred, then E would not have occurred

This is a rough characterisation of Yablo's preliminary suggestion in 'Outline for a Sketch of a Proto-theory of Causation' [2001]. The problem with it, as Yablo notices, is that it over-generates causes. To see this, consider a revised example of BILLY AND SUZY#1:

**Example (7):** METEOR: (An example similar to Yablo's [2001]) Suppose our example in BILLY AND SUZY#1 (Example 1) again; this time, suppose there is a meteor orbiting a distant planet many light years away, and that this is causally independent of the shattering. Now, the orbiting of the meteor is spuriously analysed as a cause of the shattering as follows:

- There is some fact F (i.e. the fact that the shattering occurs only if the meteor is still on its Mars orbit) such that had it still obtained but the meteor's orbit not occurred, then the shattering would not have occurred

The problem here is that it seems we can hold any old fact fixed along with an effect in order to generate spurious causal claims. We note here that where sentence constructions such as 'the shattering occurs only if the meteor is still on its Mars orbit' express a fact, it certainly does not express an event as it was specified in section 3. Indeed, events are

only those concrete 'rock solid' things that don't include abstract logical relations which the latter expression appeals to. Hence our special appeal to events, and not facts, in our above revision to the Humean analysis.

So the holding-fixed analysis meets with a lot of success with respect to early pre-emption cases. As we shall see, this analysis will suggest some morals for a successful counterfactual analysis of causation further down the line when we come to chapter 4. For now, I turn to Lewis' later analyses.

*Lewis' (2000) influence analysis succeeds* (in all early pre-emption examples thus far): Recall the details of Lewis' influence analysis, where C and E are distinct actual events:

- C causes E iff C influences E.
- C influences E iff there is a substantial range of $C_1 \dots C_n$ of different not-too-distant alterations of C (including the actual alteration of C) and a range of $E_1 \dots E_n$ of alterations of E, at least some of which differ, such that if $C_1$ had occurred, $E_1$ would have occurred,... if $C_n$ had occurred, $E_n$ would have occurred.[83]

For good measure, I will demonstrate how this analysis deals with our Billy and Suzy cases (1) and (3), as it is then obvious how the analysis succeeds for the other early pre-emption cases we have visited thus far. As follows:

---

[83] Lewis, [2001], p. 91

- Suzy's throw (C) influences the shattering (E). Or in other words, there is a substantial range of different not-too-distant alterations of Suzy's throw (C) (i.e. $C_1$ Suzy throws earlier, $C_2$ she throws later, $C_3$ harder, $C_4$ softer… etc.) and a range of alterations of E (i.e. $E_1$ the shattering occurs earlier, $E_2$ the shattering occurs later, $E_4$ the shattering occurs more violently, $E_4$ less violently… etc.) such that the following counterfactuals are true: Suzy throws earlier ($C_1$) $\square\rightarrow$ the shattering occurs earlier ($E_1$), Suzy throws later ($C_2$) $\square\rightarrow$ the shattering occurs later ($E_2$), Suzy throws harder ($C_3$) $\square\rightarrow$ the shattering occurs more violently ($E_3$), Suzy throws softer ($C_4$) $\square\rightarrow$ the shattering occurs less violently ($E_4$)… etc.

Thus, the influence analysis analyses Suzy's throw (C) as a cause of the shattering (E); the correct analysis. However, we should note here that the influence analysis seems to analyse Billy's presence as a causal influencer of the shattering (for similar reasons as the fine-grained Humean analysis did) - that is, very small alterations to Billy's presence counterfactually imply very small alterations to the shattering (here we can suppose that this is due to gravitational factors that relate Billy's presence with the shattering). Is this a problem? No. Recall in our expository section (2.4.3) that Lewis' concept of influence is designed to analyse a notion of causal *influence*, and that influence can be viewed as a comparative notion. That is, we can say:

- If the not-too-distant alterations of C counterfactually imply alterations of E that are more dramatic alterations to E than the alterations to E that the not-too-distant

alterations of A counterfactually imply, then C causally influences E more than A
does.

As a consequence of this view, we are led to the conclusion for our above example that although it is true that both Billy's presence (A) and Suzy's throw (C) are causes (or causal influencers) of the shattering, Suzy's throw (C) causally influences the shattering much *more* than Billy's presence (A) does; an intuitively correct result.

Thus, Lewis' influence analysis seems to generate intuitive results concerning our examples. As we can see, the analysis derives its strength from a certain metaphysical principle being true, which we should pin down and make explicit here:

- Close alterations of the cause counterfactually imply alterations to the effect.

Making this principle explicit, unfortunately for Lewis, now makes it easier for us to find early pre-emption counter-examples that offend this principle. We shall consider one example here that shows how the influence analysis is not *necessary* for causation via a case of early pre-emption; we shall only later consider examples where it is not *sufficient* for causation (namely when we visit cases of late pre-emption). Here is the example:

**3.1.4 Modally fragile early pre-emption**

**Example (8):** The following is an example of early pre-emption in which the cause is modally fragile (i.e., it would have failed to cause the effect had it been altered in any close way). Term such a case of early pre-emption a case of 'modally fragile early pre-emption'.

BILLY AND SUZY#4: (my own example): Consider the example of BILLY AND SUZY#2 (Example 3) again; Recall, the essential ingredient of the example that made it so vicious was the fact that had Suzy not thrown her rock, Billy's smartrock would have caused a precise fragile duplication of the shattering that Suzy actually caused. Now we toughen up the example; we still suppose that Billy's smartrock would have caused the fragile version of the shattering had Suzy not thrown, but we also suppose the additional contrivance that Billy's rock is a *super*-smart rock - *had Suzy made her throw in any different way*, *or had it flown in any different way than the way it actually did*, Billy's rock would have destroyed Suzy's rock with its laser guns (it now has laser guns by the way…), and then exploded (leaving the window undamaged).

**Example (9):** Another case of 'modally-fragile early pre-emption'.

CAR RACE#3 (a variation of an example also due to Kvart[84]): Suppose the example of CAR RACE#1 (example 2) with some additional details. Suppose *the wagon could not have started in any other time than the way it actually did*; so suppose the organisers of the race are very strict about when the wagon can start the race. Also suppose that *the*

---

[84] Kvart, [2000], p. 411

*wagon would have immediately broken down and failed to start if its starting had occurred in any different but similar manner*; so suppose the wagon is one of those extremely temperamental vehicles that just won't work unless it is started very carefully and in a very particular manner. Suppose also, that in keeping with the rules of the race, both drivers could only drive at one speed throughout, that both vehicles must start the race at the same time, and must only drive in vehicles which have precise specifications (including colour, structure, wheel size etc.), and if they break these rules they get immediately stopped by referees.

*The influence analysis* fails in the first example:

- It is *not* the case that Suzy's throw (C) influences the shattering (E). Or in other words, it is *not* the case that there is a substantial range of different not-too-distant alterations of Suzy's throw (C) (i.e. $C_1$ Suzy throws earlier, $C_2$ she throws later, $C_3$ harder, $C_4$ softer… etc.) which counterfactually implies a range of alterations $(E_1, E_2, E_3, E_4)$ of the shattering (E). Indeed, $\sim(oC_1 \,\square\!\!\rightarrow oE_1)$, $\sim(oC_2 \,\square\!\!\rightarrow oE_2)$, $\sim(oC_3 \,\square\!\!\rightarrow oE_3)$, $\sim(oC_4 \,\square\!\!\rightarrow oE_4)$… etc.

We achieve the latter list of counterfactuals from the observation that any close alteration of the throw would have resulted in the same shattering (E), because the smartrock would have consistently caused the same version of the shattering in any close world in which Suzy's throw had been altered. Thus there is no 'substantial' range of not-too-distant alterations of the shattering counterfactually implied by close alterations of Suzy's throw,

and thereby we are forced to the conclusion that Suzy's throw (C) did not influence/cause the shattering (E). The influence analysis obviously fails in the second example upon very similar reasoning. Might the influence chains analysis come to the rescue here with respect to these two examples? No: as follows;

*Lewis' influence-chains analysis fails in Car Race#3*: Recall the details of Lewis' influence chains analysis:

- C causally influences E iff there is a chain of influence from E to C.
- There is a chain of influence from E to C iff there is a set of actual events E, D, $D_1$ etc… C, such that C influences…, $D_1$ influences D, etc…. E.

Unfortunately, this analysis (although as successful as the influence analysis so far in analysing cases of early pre-emption), won't help us here; we have set up our examples (6) and (7), such that if any event in the process from (C) to (E) had occurred, the pre-empted cause would have caused the same fragile version of the effect in the actual world. This raises the question; if so far the influence and influence-chains analyses have been demonstrated to be just as good as each other, why bother with the latter revision? The answer comes when we visit cases of early pre-emption which the influence analysis cannot analyse but the influence chains analysis can. It will be instructive to visit such a case to demonstrate the full superiority of the influence chains analysis, as follows:

**Example (10):**

BILLY AND SUZY#5 (my own example): Suppose our smartrock example again (BILLY AND SUZY#2, example 3), this time, suppose only that if Suzy had altered her throw, Billy's smartrock would have destroyed her rock just before it hit the window, and the smartrock would have caused an exact duplication of the actual shattering. Suppose also that Billy's smartrock is insensitive to how Suzy's rock actually flies, and is only sensitive to the physical movements that Suzy makes when actually making the throw.

It is easy to see how the influence analysis fails - close alterations to Suzy's throw (C), do not counterfactually imply alterations in the shattering (E). However, close alterations to Suzy's throw do counterfactually imply a range of alterations to her rock in mid-trajectory (D), because the smartrock would only have destroyed Suzy's rock when it reached near the window. In addition, alterations to Suzy's rock in mid-trajectory counterfactually imply a range of alterations to the shattering, given that in such counterfactual scenarios Suzy's throw would have been the same as it was in the actual world and Billy's smartrock is insensitive to how Suzy's rock flies in mid-flight. Thus, (C) influences (D), and (D) influences (E), and thus C is analysed as a cause of E; the correct analysis. Thus the influence chains analysis succeeds here where the influence analysis fails.

However, the benefits of the influence chains analysis, as we have seen, are not enough to help us in cases of modally fragile early pre-emption (examples (6) and (7)). Thus we must continue our investigation for analyses that can handle these examples adequately.

Let us continue in this vein. We noted earlier that a 'holding fixed' strategy paid off with regard to cases of early pre-emption: might such a strategy help us with the influence analysis? The answer is that it does, as follows:

*Lewis' holding fixed influence analysis succeeds*: (my own revision to the influence analysis). There is an intuitive way to revise Lewis' influence analysis such that it gives us the correct result in the cases we have visited so far; the strategy comes from the idea of 'holding events fixed' which we saw earlier viz. the holding fixed analysis. First I'll provide the detailed revisions we need for the notion of influence-fixing, then we'll see how it succeeds. The revision is given as follows; where C, E and F are distinct actual events, then

- C causes E iff C F-influences E,
- C F-influences E iff there is a substantial range of $C_1$… Cn of different not-too-distant alterations of C (including the actual alteration of C) and a range of $E_1$… En of alterations of E, at least some of which differ, and some event F such that if $C_1$ & F had occurred, $E_1$ would have occurred,… if $C_n$ & F occurred, $E_n$ would have occurred.

Term this analysis the 'holding fixed influence analysis', given it employs a similar strategy to the holding fixed analysis. Return to the example of SUZY AND BILLY#4 (example (8)). We succeed in analysing the example as follows:

- Suzy's throw (C) F-influences the shattering (E). Or in other words, there is a substantial range of different not-too-distant alterations of Suzy's throw (C) (i.e. ($C_1$) Suzy throws earlier, ($C_2$) she throws later, ($C_3$) harder, ($C_4$) softer… etc.) and a range of alterations of E (i.e. ($E_1$) the shattering occurs earlier, ($E_2$) the shattering occurs later, ($E_3$) the shattering occurs more violently, ($E_4$) less violently… etc.) such that (and where F is the event of Billy's supersmart rock lying inertly in his hand throughout) $(((oC_1 \mathbin{\&} oF) \mathbin{\square\!\!\rightarrow} oE_1)$, $((oC_2 \mathbin{\&} oF) \mathbin{\square\!\!\rightarrow} oE_2)$, $((oC_3 \mathbin{\&} oF) \mathbin{\square\!\!\rightarrow} oE_3)$, $((oC_4 \mathbin{\&} oF) \mathbin{\square\!\!\rightarrow} oE_4)$… etc.)

The success of the analysis here relies on the fact that we can hold fixed the event in which Billy holds onto his super-smart rock throughout. If we hold fixed this event in our counterfactuals, then we are free to take advantage of a whole range of alterations of Suzy's throw (C) that in turn results in alterations to the shattering (E). The influence fixing analysis succeeds in example (9) upon similar reasoning, as follows:

- The wagon starting off on its route (C) F-influences the breaking of the finishing ribbon (E). Or, in other words, there is a substantial range of different not-too-distant alterations of the wagon-start (C) (i.e. $C_1$ the wagon starts earlier, $C_2$ it starts later, $C_3$ much earlier, $C_4$ much later… etc.) and a range of alterations of E (i.e. $E_1$ the ribbon breaking occurs earlier, $E_2$ later, $E_3$ much earlier, $E_4$ much later… etc.) such that (and where F is the event in which the fast car comes to a halt before the barrier) $((oC_1 \mathbin{\&} oF) \mathbin{\square\!\!\rightarrow} oE_1)$, $((oC_2 \mathbin{\&} oF) \mathbin{\square\!\!\rightarrow} oE_2)$, $((oC_3 \mathbin{\&} oF) \mathbin{\square\!\!\rightarrow} oE_3)$, $((oC_4 \mathbin{\&} oF) \mathbin{\square\!\!\rightarrow} oE_4)$… etc.

The success of the analysis here relies on the fact that we can hold fixed the events in which the fast car has to brake to a halt before the barrier. If we hold fixed these events in our counterfactuals, then we are free to take advantage of a whole range of alterations of the wagon-starting (C) that in turn results in alterations to the ribbon breaking (E).

The holding fixed strategy has proved a good modification for the influence and Humean counterfactual analyses alike - it seems that all the causes of early pre-emption cases we have met so far have been correctly analysed as causes by the aforementioned analyses. However, it is an unfortunate feature of any version of the influence analysis that it over-generates causes (i.e. analyses non-causes as causes). To see how this is the case, we first observe that it is a condition of (any version of) the influence analysis' sufficiency that the following general metaphysical principle is true (and stands as a close counter-part to the one we visited earlier):

- Close alterations of non-causes do not counterfactually imply alterations to the effect.

Now we have pinned down this principle, all we need to do is find an example in which close alterations of non-causes counterfactually imply alterations to the effect in a substantial way, to demonstrate that the influence analyses are not sufficient for causation. Here is one such revised CAR RACE example:

**3.1.5 Over-generative early pre-emption**

**Example (11):** The following example demonstrates how versions of the influence analysis that we have visited thus far are not sufficient for causation: call such a type of early pre-emption a case of 'over-generative early pre-emption';

CAR RACE#4 (my own example): Suppose similar details to that of CAR RACE#3 (example 9), but this time, suppose that the sports car (now blue, which didn't cause the ribbon breaking (E)), is powered by a very sensitive sort of rocket fuel. Suppose also that had the blue car setting off on its route (A) been altered in any close fashion, then the car would have caused a leak of rocket fuel, which would have caused it to race at extremely high but erratic speeds, destroying both the barrier and the red car, and passing through the finishing line first, causing the breaking of the ribbon (E).

We can easily see how any of our versions of the influence analysis would incorrectly analyse the blue car starting off on its route (A), as a cause of the ribbon-breaking (E) - any close alteration of the red car starting off on its route counterfactually implies a range of alterations to the effect, and this is true no matter what we hold fixed from the actual case. The problem becomes more perspicuous when we combine similar details about the pre-empted race car with the details of example (9); suppose the rocket fuelled car races against the fragile wagon, and that the latter causes the ribbon breaking and not the former- we now have a case in which the non cause (the rocket fuelled car) is analysed as a cause of the ribbon breaking, and the cause (the fragile wagon) is analysed as a cause.

In such a case, the influence and influence chains analysis would be demonstrated to be neither necessary nor sufficient for causation; and in one single example of early pre-emption! This is a *dramatic* result for the influence analysis; we have now shown that *the influence analysis cannot deal with some of the most basic sorts of early pre-emption!*

We should now move on to see how Lewis' [1986] quasi-dependence analysis fares up against the examples we have met so far, before making some general observations in this section:

*Lewis' quasi-dependence analysis succeeds in all cases so far*: Recall the details of Lewis' quasi-dependence analysis, where C and E are distinct actual events, then

- C causes E iff E quasi-depends on C.

- E quasi-depends on C iff there is an isonomic duplicate of a process from E to C, in which the duplicate of E counterfactually depends on the duplicate of C.

The analysis succeeds in analysing the early pre-emption cases thus far. We first show how it deals with the case of BILLY AND SUZY#1 and SMART-ROCK:

- The shattering (E) quasi-depends on Suzy's throw (C)

- There is an isonomic duplicate of a process from the shattering (E) to Suzy's throw (C) (namely, the process that constitutes the ball in flight from Suzy's hand

to the window), in which the shattering (E) counterfactually depends on Suzy's throw (C).

In short, Suzy's throw (C) causes the shattering (E) as there is such an isonomic duplicate of a process which features dependence between the shattering (E) and Suzy's throw (C); i.e. there is a possible case, very similar to that of the actual case but without Billy or his smartrock (A), in which the shattering (E) depends on the throw (C). A similar analysis holds for the CAR RACE. As follows:

- The breaking of the finishing ribbon (E) quasi-depends on the wagon starting off on its route (C)

- There is an isonomic duplicate of a process from the ribbon breaking (E) to the wagon starting off on its route (C) (namely, the process that constitutes the travelling of the wagon to the ribbon), in which the ribbon breaking (E) counterfactually depends on the wagon starting off on its route (C)..

In short, the wagon starting off on its route (C) causes the ribbon breaking (E) as there is such an isonomic duplicate which features dependence between the ribbon breaking (E) and the wagon starting (C); i.e. there is a possible case, very similar to the actual case but without the wagon's racing opponent (A) being present, in which the ribbon breaking (E) depends on the wagon-starting (C).

The quasi-dependence analysis finds its strength in its appeal to processes, and in isonomic duplicates which stand as 'similar copies' of parts of the original examples. This sort of strategy shall meet problems in other sorts of pre-emption cases (and as it turns out, these later sorts of problems can be easily adapted to cases of early pre-emption). I shall not meet them here because their full force only becomes more intuitively apparent in other examples (and in addition because we do not have enough space to visit even more cases of early pre-emption!). I now visit how Collins' [2000] would be-dependence analysis deals with our cases of early pre-emption.

*The would-be dependence analysis succeeds*: Earlier on we mentioned Collins' 'would be dependence' analysis, as an alternative development of the essential strategy of the quasi-dependence analysis. The essential strategy is to 'get rid' of background pre-empted causes that prevent dependence of the effect on the cause, and then test for dependence, and in such a way (hopefully) achieve a correct analysis. Recall some details of the 'would be dependence' analysis, where C, E and P are distinct actual events, then:

- C causes E iff there is a chain of counterfactual dependence linking C to E, or there would be such a chain had it not been for some pure dependence preventer.

- An event P purely prevents dependence of E on C iff P is a dependence preventer of a chain of events E on C, and P prevents no event in the chain of events that would form a counterfactual chain linking E to C had P not occurred.

- P is a dependence preventer of E on C iff had P not occurred and the C-E chain of events still had, then E would counterfactually depend on C.

The analysis treats the Billy and Suzy cases as follows:

- There would be a chain of counterfactual dependence linking Suzy's throw (C), to the shattering (E), had it not been for some pure dependence preventer (Billy's presence (A)).
- Billy's presence prevents dependence of the shattering (E) on Suzy's throw (C), because had it not been the case that Billy was present, the shattering (E) would have depended on Suzy's throw (C).

However, the success of this strategy meets with an immediate problem. To see this, consider a revised Billy and Suzy example, as follows:

## 3.1.6 Shunting early pre-emption

**Example (12):** An example in which it's the case that a cause of the pre-empted back-up process also causes the effect via the actual process, such that if it didn't cause the effect by the actual process, it would have caused it by another. Term this type of pre-emption 'shunting pre-emption', given the back-up cause would have shunted from one process to another.

SUZY AND BILLY#6: (my own example) Suppose the details of SUZY AND BILLY#1 (example 1), but with the following modifications: Suzy only throws her rock because Billy tells her to. Thus, Billy's presence (A) causes Suzy's throw (C), which causes her rock to fly in mid-trajectory (D), which in turn causes the shattering (E). Had Suzy not thrown her rock however, the shattering would still have occurred (E), because Billy would have thrown a rock at the window had Suzy not.[85]

*The would be-dependence analysis fails, as follows*:

- It is not the case that there would be a chain of counterfactual dependence linking Suzy's throw (C), to the shattering (E), had it not been for some pure dependence preventer (Billy's presence (A)- the only dependence preventer present). This is because had Billy not been present Suzy would not have thrown (she only throws because Billy tells her to!), and thus there would not be a chain of dependence between Suzy's throw and the shattering, because there simply would not have been a chain of events from her throw to the shattering (E).

- Billy's presence purely prevents dependence of the shattering (E) on Suzy's throw (C), because had it not been the case that Billy was present, the shattering (E) would have depended on Suzy's throw (C).

---

[85] Here we suppose it is Billy's presence at some time which is the only pure dependence preventer of the shattering (E) on Suzy's throw (C).

The problem here is that sometimes getting rid of the pure dependence preventer gets rid of the chain of events that we wished to analyse as a causal chain in the first place! Not to worry: we can simply revise the would-be dependence analysis to succeed in this case by stating in the analysis that we 'hold fixed' the relevant chain, as follows:

- C causes E iff there is a chain of counterfactual dependence linking C to E, or there would be such a chain had it not been for some pure dependence preventer, *and the events of the would-be chain still occurred*.

We shall give some additional details about a revised notion of dependence prevention later when we visit cases of late pre-emption; for now we note that the above seems a good first step in revising Collins' analysis. I conclude this section with some general observations:

*Some observations*: So far, the basic counterfactual analysis has seen two revisions; 'holding fixed' and appealing to chains. The influence analysis has also seen these revisions; that is 'holding fixed', and appealing to chains. In both cases, the holding fixed strategy has met with better successes than appealing to chains - and with the additional attraction of not having to buy wholesale into the notion that causation is a transitive relation. We also visited both the quasi-dependence analysis, with its strategy of 'getting rid of' some of the background events that made these examples so problematic, and the would-be dependence analysis, along with one revision. We met 6 types of early pre-emption examples in this section:

1. Coarse-grained early pre-emption (examples (1) and (2))

2. Fine-grained early pre-emption (examples (3) and (4))

3. Modally fragile early pre-emption (examples (8), (9), (10))

4. Non local early pre-emption (example (6))

5. Over-generative early pre-emption (example (11))

6. Shunting early pre-emption (example (12))


Our analyses and revised analyses met these cases with varying success. We have not had space to make explicit how every analysis deals with every case or type of pre-emption, but we can catalogue some of the demonstrated (and the obviously expected) results here of how each analysis deals with each of the types of early pre-emption just mentioned:


- The Humean counterfactual analysis fails in all the examples

- The holding fixed analysis succeeds in all the examples

- Lewis' counterfactual chains analysis only fails in 6

- Lewis' influence analysis succeeds in all except 8, 9, 10 & 11

- Lewis' influence chains analysis succeeds in all except 8, 9, & 11

- The holding fixed influence analysis succeed in all examples except 11

- Lewis' quasi-dependence analysis succeeds in all the examples

- The would-be dependence analysis succeeds in all cases except 12

- The revised would-be dependence analysis succeeds in all the examples

Three of the analyses we considered so far have met with good results. The 'holding fixed strategy' has proven especially effective as a revision for Hume's basic analysis, and the 'getting rid of strategy' has proven effective for the quasi-dependence analysis and revised would-be dependence analysis. Why have these two strategies prospered so well? The answer comes in their treatment of the troublesome pre-empted back-up causes - in either case, the troublesome pre-empted back-up has either been 'held fixed' or 'got rid of' in such a way that it no longer interferes with the effect counterfactually depending on the cause in the scenarios required for a successful analysis.

I shall now turn to cases of 'late pre-emption'. Cases of late pre-emption differ subtly from cases of early pre-emption, but as we shall see, these subtle differences make all the difference with respect to how successful some of our analyses are.
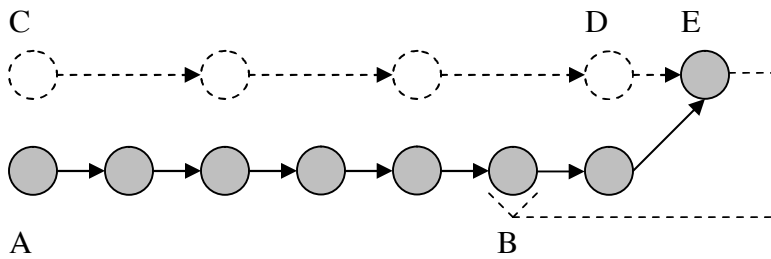
## 3.2 Late pre-emption

*Intuitive explanation*: In cases of late pre-emption, the causal process running from the pre-empted cause is cut short from causing the effect only *after* the main causal process has gone to completion and caused the effect.

*Diagrammatic reconstruction*: Consider the diagrams below: Events C and A occur simultaneously; C causes E (via D) just before A is able to cause E; such that C pre-empts the process from A to E from coming into completion. C is the pre-empting actual cause of E, and A is the pre-empted potential back-up cause.
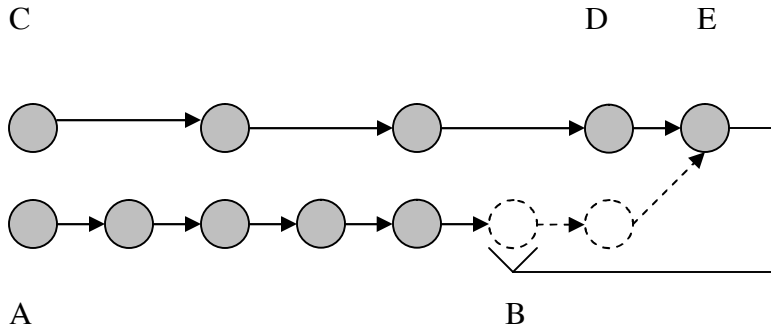
The actual world: C causes E via D, the fact that D causes E inhibits B from causing E



The closest possible world in which C does not occur: A causes E via B

## 3.2.1 Coarse-grained late pre-emption

**Example (13)**: An example of late pre-emption in which only a coarse-grained version of the effect still would have occurred had the actual cause not occurred. Call such an example an example of 'coarse-grained late pre-emption'.

SUZY AND BILLY#1* (due to Lewis[86]): Suppose Billy and Suzy are playing with rocks again. Suppose Suzy throws a rock (C) at a window, causing the rock to fly just next to

---

[86] Lewis, [2000]

the window (D), in turn causing the shattering of the window (E), such that Suzy's throw (C) causes the shattering (E). Suppose that Billy throws a rock (A) towards the window at the same time as Suzy throws hers (C), such that Billy's rock flies close to the window (B). But suppose Suzy's rock managed to cause the shattering (E), just moments before Billy's rock was able to, such that Billy's rock passes through the empty window frame and to the ground (without ever touching the window with impact). Thus, Suzy's throw (C) causes the shattering (E), and not Billy's throw (A).

**Example (14):** Another example of 'coarse-grained late pre-emption'.

CAR RACE#1*: (due to Kvart[87]) Suppose that a sports car and a wagon are starting on a race towards the finishing line from a shared start point. Suppose the wagon starts on its route (C), causing it to travel just before the finishing line (D), in turn causing it to break the finishing ribbon (E), such that the wagon-starting (C) causes the ribbon breaking (E). Suppose that the sports car starts off on its route (A) at the same time as the wagon starts its route (C), such that the sports car travels close towards the finish ribbon (B). But suppose that the wagon only managed to beat the sports car in breaking the finishing ribbon (E), just moments before the sports car was able to, such that the sports car passes through the finishing line without breaking the ribbon. Thus, the wagon-starting (C) causes the ribbon-breaking (E), and not the sports car starting (A).

---

i) *The basic Humean analysis fails*: Recall the details of the basic Humean analysis; where C and E are distinct actual events, then:

- C causes E iff E counterfactually depends on C.

The basic Humean analysis fails to successfully analyse either example. We should be suitably comfortable with the details of how this is so, without having to go into the precise details of the Lewisian semantics. Suffice to say the analysis fails in each examples as follows:

- The window shattering (E), does not counterfactually depend on Suzy's throw (C), because if Suzy's throw (C) had not occurred, Billy's throw (A) would have caused a similar (albeit slightly later) shattering (E).
- The ribbon-breaking (E), does not counterfactually depend on the wagon starting (C), because if the wagon starting (C) had not occurred, the sports car's starting (A) would have caused a similar (albeit slightly later) ribbon breaking (E).

Might a 'fine-grained Humean analysis', which we visited in the section on early pre-emption, prove better in such examples? Recall the details:

- C causes E iff E counterfactually depends on C, and C and E are distinct actual events (but E is a fine-grained event).

For the above two examples it is obviously true that the fine-grained Humean analysis succeeds; for in the first example it is true that if Suzy's throw had not occurred, the fine-grained version of the window shattering would not have occurred (given only a different shattering would have occurred slightly later courtesy of Billy's throw). For the second example, it's true that if the wagon had not started off on its route, the fine-grained version of the ribbon breaking would not have occurred (given only a different version of the ribbon breaking would have occurred slightly later courtesy of the sports car).

It should be obvious by now (from what we can learn from our examples in the section on early pre-emption) that we can provide revisions to our late pre-emption examples such that the pre-empted back-up cause would have caused an exact same fragile version of the shattering. We state the revisions to the examples here for good measure, where both stand as examples of 'fine-grained late pre-emption'.

### 3.2.2 Fine-grained late pre-emption

**Example (15)**:

BILLY AND SUZY#2* (an example suggested by Yablo[88]): Suppose the details of BILLY AND SUZY#1* (example 13) again, except this time, suppose Billy has his smartrock, fully equipped with a laser guidance system, rocket boosters, and on-board computer, which would have enabled and determined the smartrock to cause the exact

---

[88] See Collins, Hall, and Paul [2004]

same version of the shattering that Suzy's rock actually did, had Suzy not made her throw (suppose it would have taken the exact same trajectory that Suzy's rock actually did, and upon certain and complicated accelerations/decelerations etc. entered the window at exactly the same time and manner as Suzy's rock actually did).

**Example (16):**

CAR RACE#2* (see Kvart[89]): Suppose the details of CAR RACE#1 again (example 2), except this time, suppose that the driver of the sports car was holding back from accelerating to faster velocities throughout the race (suppose this is because the sports car driver was scared of crashing into the wagon which always had a very slight lead on the sports car). Suppose, however, that if the wagon had not started the race, the sports car driver would have had more confidence to go slightly faster (given the driver would feel more comfortable driving at higher speeds without other dangerous and local locomotives), and, as luck would have it, the ribbon breaking that would have occurred upon these circumstances would have been the exact same fragile version of the ribbon-breaking that the wagon caused in the actual case.

The first example here is pretty contrived (smartrock examples seem to be), but the second is far more realistic (in fact it seems a pretty ordinary example). In any case, they demonstrate an important point; there are possible cases (no matter how peculiar) that

---

[89] Kvart, [2000], p. 411

escape successful analysis by attempts to 'fine-grain' the effect, and thus that the fine-grained Humean analysis is not necessary for causation overall.

I now assess whether the 'dependence-fixing analysis' has any better luck in dealing with late pre-emption cases.

*The dependence-fixing analysis fails*: Recall the details of the dependence-fixing analysis: Where C, E, and F are distinct actual events, then:

- C causes E iff there is some event/s F such that if F had still occurred and C not, then E would not have occurred.

The analysis fails in our late pre-emption examples. In each of our cases, there are no events that could be held fixed such the closest not-C worlds are not-E worlds. In any of our late pre-emption examples, a version of E still would have occurred had for any F, some F still occurred and C not. Thus, the dependence fixing analysis fails to correctly analyse C as the cause of E in our examples.

But all is not lost here. We can appeal to a fragile version of E in our analysis, and the analysis yields the correct results. To see this, we need but consider how the analysis deals with example (13):

- There is some F (namely, the event of the smartrock travelling only just behind Suzy's rock) such that if F had still occurred and Suzy's throw (C) not occurred, the fragile version of the shattering (E) would not have occurred.

Of course, the analysis shares the same initial problems as our fine-grained Humean analysis did. That is, the analysis suffers from incompleteness - it cannot analyse causation between coarse-grained events. We move on to the influence analysis:

*The influence analysis succeeds*: Recall the details of the influence analysis; where C and E are distinct actual events, then:

- C causes E iff C influences E.
- C influences E iff there is a substantial range of $C_1 \ldots C_n$ of different not-too-distant alterations of C (including the actual alteration of C) and a range of $E_1 \ldots En$ of alterations of E, at least some of which differ, such that if $C_1$ had occurred, $E_1$ would have occurred,... if $C_n$ had occurred, $E_n$ would have occurred.[90]

There is a worry about how the influence analysis might succeed in our four late pre-emption examples. To see this, we notice that not only is it the case that close alterations to Suzy's throw (C) counterfactually imply alterations to the shattering (E), but also that close alterations to Billy's throw (A) counterfactually imply alterations to the shattering

---

[90] Lewis, [2003], p. 91

(E) (and a substantial range in both cases). Thus, prima facie, it seems as if the influence analysis analyses *both* Suzy's and Billy's throw as a cause of the shattering; the incorrect analysis.

One way around this result is to appeal to the notion of 'a relatively inert zone' that we visited in our clarificatory sections on the influence analysis. To see how such a clarification works in the influence analysis' favour, we notice that there is a very close 'zone' of worlds in which alterations to Billy's throw imply alterations to the shattering (a 'relatively inert zone'). We then note that in the *comparatively* close set of worlds in which alterations of Suzy's throw occur, these latter alterations do indeed imply alterations to the shattering. Thus, we achieve our analysis that Suzy's throw caused the shattering but Billy's didn't because there is a close zone of worlds in which Suzy-alterations counterfactually imply shattering alterations, but in that relatively close zone of worlds Billy's throw alterations are 'relatively inert' with respect to alterations of the shattering. The strategy in achieving the correct causal analysis is then to pick out a 'relatively inert zone' in which alterations to non causes do not imply alterations to an effect, where in that zone alterations to actual causes do.

Unfortunately, this modification won't help us when we consider late pre-emption examples in which either:

- Close alterations to the cause do not counterfactually imply alterations to the effect

- Close alterations to a non cause counterfactually imply alterations to an effect in *all* not-too-distant worlds.

We visited these sorts of cases when we treated cases of early pre-emption. In the former cases, we demonstrate that the influence and influence chains analyses alike are unnecessary for causation; in the latter we demonstrate that both are insufficient. We give one example of the former case (which shall be termed 'modally fragile late pre-emption') and one of the latter (termed 'over-generative late pre-emption'). We shall not spend too much time on them or how the analysis deals with them, as the problems that they pose for the influence analysis are pretty much the same as the problems we visited in the section on early pre-emption.

### 3.2.3 Modally fragile late pre-emption

**Example (17):**

CAR RACE#3*: (my own example) Suppose the details of CARRACE#2* again (example 16), except this time, suppose the traveling racing wagon is exceptionally fragile, and would have immediately fallen apart and been disqualified from the race, had it been alerted in any fashion.

In this example, we can see that it is not the case that not-too-distant alterations to the cause (the route of the wagon), counterfactually imply alterations to the effect (the

breaking of the ribbon), because had alterations to the cause been made, it would have rendered it causally inefficacious (the wagon would have broken down), and the ribbon breaking would have been caused in the same fashion over the not-too-distant worlds by the race car. Thus, the cause is not analysed as a causal influencer of the effect; the incorrect analysis.

### 3.2.4 Over-generative late pre-emption

**Example (18):**

CAR RACE#4*: (my own example) Suppose example CAR RACE#3* (example 17) again, except this time suppose that the race car driver (who loses the race) is in cahoots with evil scientists. Suppose the driver has been instructed by the evil scientists to drive at exactly 70mph during the first half of the race, but accelerate to over that speed half way through the race. Suppose that if the race car driver accelerated over or dipped below 70mph, a pair of rocket boosters would activate causing the race car to win the race at outrageously high velocities. Suppose that the driver's car has a malfunction after the first ¼ of the race, resulting in the fact that the race car can't shift off 70mph. The race car loses the race (only just - by a microscopic distance in fact), and the wagon wins by breaking the ribbon (E).

In this example, we can see that not-too-distant alterations to a non cause of the ribbon breaking (the race car driving over half way through the race), counterfactually imply a

substantial range of alterations to an effect (the ribbon breaking), because had the race car at such points been altered in its velocity, then it would have accelerated to superluminal velocities which would have therein altered the ribbon breaking. Thus, a non-cause is analysed as a causal influencer of the effect; the incorrect analysis.

Examples (17) and (18) thus serve to demonstrate how the influence analysis is neither necessary nor sufficient for causation. We can easily see that appealing to modifications to the influence analysis (such as that provided by the influence chains or influence fixing analysis, or by appealing to 'relatively inert zones') do not help us here in either of these examples, and so our influence analyses are neither necessary nor sufficient for causal influence.

So far our examples of late pre-emption have been only slightly tougher examples for our analyses. The real trouble with these examples confronts the quasi-dependence and would be dependence analyses, which (upon certain interpretations) appear to over-generate causes by appeal to any of the late pre-emption examples visited thus far:

*The quasi-dependence analysis has mixed results:* Recall the details of the quasi-dependence analysis; where C and E are distinct actual events, then:

- C causes E iff E quasi-depends on C.
- E quasi-depends on C iff there is an isonomic duplicate of a process from E to C, in which the duplicate of E counterfactually depends on the duplicate of C.

First we notice that in all of our examples there is an isonomic duplicate containing C, E and the process running from C to E which elicits counterfactual dependence. Take any of the CAR RACE late pre-emption examples: There is an isonomic duplicate of the wagon starting, and the process of the wagon travelling from the start point to the ribbon, in which the ribbon breaking (E) counterfactually depends on the wagon starting (C). There is such a duplicate, because there are possible cases (just like the actual one in which the wagon starts, travels, and breaks the ribbon), in which certain background events are not present (such as the sports car travelling on its route), and in which the relevant sort of dependence holds. Thus the analysis analyses the cause as a cause in each of the examples, and demonstrates an early success.

Whether the quasi-dependence analysis analyses the sports car's route as a non-cause is contingent on what we take isonomic duplicates to be. Recall, Lewis' criterion was that isonomic duplicates didn't need to be precise, strict duplicates of events, but simply reasonably similar ones. Taking Lewis at his word, his analysis over-generates causes by analysing the pre-empted back-up cause as an actual cause. To see this, we need but consider any of the cases of CAR RACE, as follows:

We suppose that in the actual case in which the sports car takes its route, the car is a microscopic distance away from breaking the ribbon, but was just caught short by the wagon. Now, there is an isonomic duplicate of the process of the sports car travelling on its route up to the ribbon, in which the ribbon breaking (E) counterfactually depends on

the wagon taking its route (C). This is the case for two reasons: firstly this is because Lewis' notion of a duplicate (i.e. as a very similar case indeed) is loose enough to permit of counterpart cases in which the sports car really does cause the ribbon breaking (the actual case in which it doesn't and the possible case in which it does are only microscopically different), and secondly, because there are such very similar possible cases in which the sports car does cause the ribbon breaking and counterfactual dependence is elicited between the ribbon breaking and the sports car travelling on its route (we need only think of counterpart cases in which there is no wagon). Thus, Lewis' quasi-dependence analysis analyses the pre-empted back-up cause as a cause in this example, and is thus *insufficient* for causation.

A possible revision might help Lewis here, as follows: We might revise the quasi-dependence analysis such that the isonomic duplicates appealed to in our analysis are strict isonomic duplicates (i.e. cases with exactly the same events and laws). However, the problem with this idea is that the quasi-dependence analysis (call the revised version the 'strict' quasi-dependence analysis) would be unnecessary for causation (i.e. it would not always analyse all causes as causes). To see this, consider the following example of late pre-emption:

### 3.2.5 Strengthened late pre-emption

126

**Example (19)** SUZY AND BILLY#4* (a case of my own): A case in which the cause causes the pre-empted back-up cause; call such a case a case of 'strengthened late pre-emption'.

Suppose Billy and Suzy both throw rocks at a window in a world in which strong determinism is true (see introduction for the overview of the notion of strong determinism); Suzy throws her rock (C), and smashes the window (E), just before Billy's throw (A) was able to. Suppose, also, that Billy's throw was sufficiently determined, given the deterministic laws, by Suzy's throw (so suppose Billy is very impressionable, and always participates in those mischievous acts which his sister does). Suppose also that Billy's throw was independently and sufficiently determined, given the laws, by his mischievous nature, such that if Suzy hadn't thrown her rock, he would have thrown his rock anyway. We suppose (in this strongly deterministic example) that the following laws are true; i) Whenever Suzy throws a rock at a window, so does Billy, ii) Whenever Suzy or Billy throws a rock, there is a window shattering.

*The strict quasi-dependence analysis fails*: It is easy to see how the strict quasi-dependence analysis fails to analyse the cause of the shattering (E) as Suzy's throw (C). This is because any isonomic duplicate of C and E, and the process running from C to E, will not elicit counterfactual dependence, given any isonomic duplicate must also contain the pre-empted back-up (A). This is because Suzy's throw (C) *nomologically determines* the presence of the pre-empted back-up cause (A) of the shattering; Billy's throw (A), and thus the shattering (E), will not counterfactually depend on (C) in such an isonomic duplicate.

The general problem then is that if we insist on our isonomic duplicates having *exactly* the same laws and events with respect to what the isonomic duplicate is an isonomic duplicate of, and there are cases in which a cause of an effect nomologically determines the presence of pre-empted back-up causes, then in such cases there will be no way to analyse such causes as causes of the effect given the effect won't be counterfactually dependent on the cause in any isonomic duplicate.

We are now in a dilemma - if we accept the loose Lewisian criteria for isonomic duplication then we face the problem of 'all too similar pre-emption', which generates spurious causes by appeal to all too similar cases in which causation actually holds. On the other hand, if we accept the strict criteria for isonomic duplication then we face the problem of 'background nomological determiners', where causes are not analysed as causes by the analysis in cases in which the cause nomologically determines the presence of background pre-empted causes. Is there a way out of the dilemma for Lewis?

*Solution*: Maybe one way to avoid the dilemma would be to tighten the constraint of what events may feature in isonomic duplicates (strict duplicate events), but loosen what laws we allow to feature (i.e. we say 'similar enough laws') - but this has the dangerous risk of over-generating spurious causes, given it might just take a minor revision of some nomological relations to generate new causal relations; so we should preferably avoid this strategy.

A better solution might be to analyse isonomic duplication as follows:

- X is an isonomic duplicate of Y iff X and Y have exactly the same events, and the same *governing* laws.

- Two duplicate sets of events X and Y have the same *governing* laws iff it's the case that events A and B are in X and A nomologically determines B iff events A' and B' are in Y and A' nomologically determines B'.

How might this proposed re-analysis of isonomic duplication circumvent the dilemma? Recall, the 'problem of all too similar pre-emption' was that giving way to anything other than strict duplication of events between isonomic duplicates had the potential to generate spurious causes - we avoid the problem by maintaining that the isonomic duplicates must indeed have strictly identical events. How does the analysis circumvent the second horn of the dilemma? Recall, the 'problem of background determiners' was located in the fact that the actual cause nomologically determined the presence of pre-empted back-up causes - we avoid this by not requiring that the isonomic duplicates have *all* the same laws, but merely the ones that maintain the structure of nomological determinacy within the two sets of events; problematic laws that determine pre-empted back-up causes outside the set of the isonomic duplicates are then not required to feature. Thus the dilemma seems to be circumvented; the solution comes from the fact that isonomic duplicates must contain strict duplicates of events, but only the same *governing* laws (i.e. the laws that are *relevant* to the causal structure of the case).

We should now visit how the revised would be dependence analysis treats cases of late pre-emption.

*The revised would-be dependence analysis succeeds*: Recall the details of the revised would-be dependence analysis; where C and E are distinct actual events, then:

- C causes E iff there is a chain of counterfactual dependence linking C to E, or there would be such a chain had it not been for some pure dependence preventer, and the would-be chain still occurred

- An event P purely prevents dependence of E on C iff P is a dependence preventer on a chain of events E on C, and P prevents no event in the chain of events that would form a counterfactual chain linking E to C had P not occurred.

- P is a dependence preventer of E on C iff had P not occurred and the C-E chain of events still had, then E would counterfactually depend on C.

The would-be dependence analysis analyses all cases of late pre-emption correctly. Consider a Billy and Suzy late pre-emption case to see this:

- There would be a chain of counterfactual dependence between actual events linking Suzy's throw (C), to the shattering (E), had it not been for some pure dependence preventer (Billy's throw (A)), and the would-be chain still occurred.

- Billy's throw purely prevents dependence of the shattering (E) on Suzy's throw (C), because had Billy's throw not occurred, and the would-be chain still had, the shattering (E) would have depended on Suzy's throw (C).

More revisions shall come for would-be chains later on. For now we note the success of the would-be dependence analysis in analysing Suzy's throw as a cause of the shattering, and also its success in not analysing Billy's throw as a cause of the shattering. Billy's throw is analysed as not a cause of the shattering because the dependence of the shattering on Billy's throw is *not* purely prevented - the dependence preventer (Suzy's throw) of the shattering on Billy's throw prevents an actual event in the would-be chain between the shattering and Billy's throw (i.e. the event in which Billy's rock actually reaches the window). We conclude this section with some general observations:

*Some observations*: We visited the following different types of late pre-emption in this section:

1. Coarse-grained late pre-emption (examples (13) and (14))

2. Fine-grained late pre-emption (examples (15) and (16))

3. Modally fragile late pre-emption (example (17))

4. Over-generative late pre-emption (example (18))

5. Strengthened late pre-emption  (example (19))

Our analyses and revised analyses met these cases with varying success. We have not had space to make explicit how every analysis deals with every case or type of pre-emption, but we can catalogue some of the demonstrated (and the obviously expected) results here of how each analysis deals with each of the types of *early* pre-emption just mentioned:

- The basic Humean counterfactual analysis fails in all the examples

- The Humean counterfactual analysis fails in all the examples

- The holding fixed analysis fails in all the examples

- The fine-grained holding fixed analysis succeeds in all examples

- Lewis' counterfactual chains analysis fails in all examples

- Lewis' influence, influence fixing, and influence chains analyses succeed in all examples except 17 and 18

- Lewis' quasi-dependence analysis succeeds in all examples except 18 & 19

- The revised quasi-dependence analysis (by appeal to 'relevant governing laws') and revised would-be dependence analyses succeed in all cases

We note here that *late* pre-emption cases are harder to analyse than early pre-emption cases overall. Nonetheless, three analyses still stand out as successful; the revised quasi/would-be dependence analyses, and the revised holding fixed analysis. Once again, it has been the analyses that appeal to strategies of 'getting rid' and 'holding fixed' that have proved the most successful. We now move on to cases of middle pre-emption.
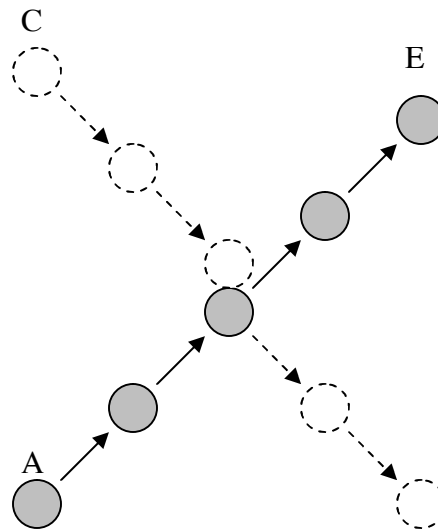
**3.3 Middle Pre-emption**

*Intuitive explanation*: Cases of 'middle pre-emption' (my terminology) are cases of causation in which some event, in the 'middle' of some causal process from C to E, prevents a back-up pre-empted process from causing the effect. The cases of 'middle preemption' were innovated and discussed by Strevens [2003].

*Diagrammatic reconstruction*: C and A jointly cause E via D. If C had not occurred, D wouldn't have, but A still would have caused E.
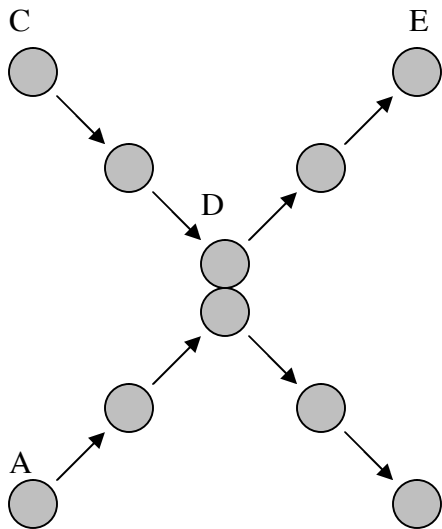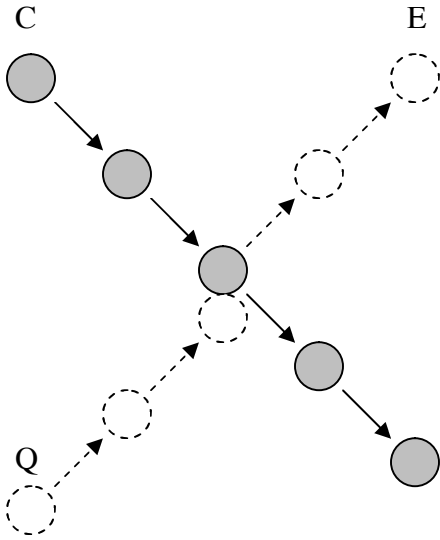
The actual world (w0)                    The closest ~C worlds to w0

The closest ~Q worlds to w0



## 3.3.1 Basic middle pre-emption

**Example (20)**:

SYLVIA AND BRUNO#1 (an example due to Strevens [2003]): Suppose Sylvia and Bruno both throw an intrinsically identical steel ball each at the same time, but that Bruno is aiming at a glass bottle, and Sylvia is aiming at a different target beyond Bruno and the bottle. Suppose Sylvia throws her ball (C) at time t-1 and her ball shatters the glass bottle (E) at t+1 after colliding with Bruno's ball (D) at t. Suppose Bruno throws his ball (A) at the same time as Sylvia throws her ball (C), but that his ball does not shatter the glass

bottle, because it collides with Sylvia's ball at a mid air collision (D) at time t, and bounces away from otherwise causing the shattering (E).[91]

This is quite a tricky example, and probably easier to understand with diagrams than with words: The following is a topographical diagram representing the trajectories of the two balls. The black line represents the trajectory of Sylvia's ball, and the dotted line represents the trajectory of Bruno's ball; the time line is plotted on an x axis below the diagram.

Sylvia's ball            Glass Bottle

Bruno's ball

T – 1     T     T + 1
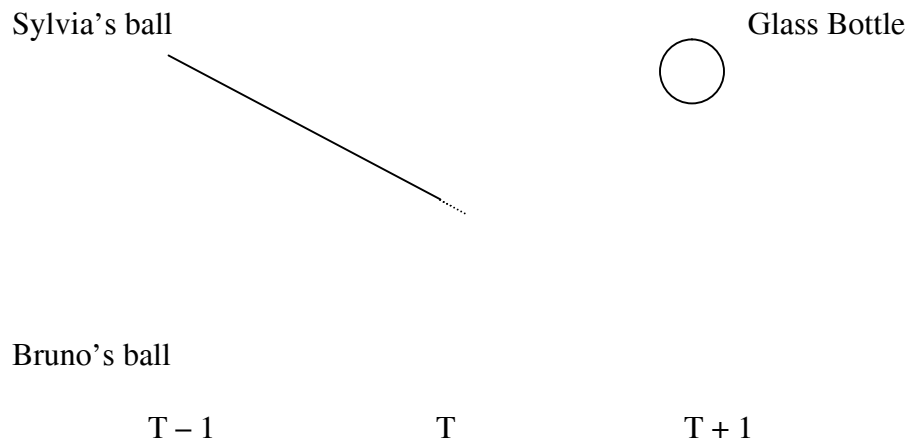
We add some further details to toughen up the example: Suppose in the range of not-too-distant possible worlds that no alteration of Sylvia's throw, or Sylvia's ball in flight towards the collision, would have resulted in an alteration of the shattering of the bottle,

---

[91] Note that the events to do with a ball flying after the collision up to the shattering would have been exactly the same had Sylvia not thrown her ball, given Bruno's ball is intrinsically identical and would have assumed the same trajectory had Sylvia not thrown.

because even the smallest alteration in Sylvia's throw, or to Sylvia's ball in flight towards the collision, would make it relatively impossible for the collision to occur, and so Bruno's ball would have taken the trajectory of Sylvia's ball after time t, and caused the shattering of the glass bottle (E) in exactly the same time and manner as it actually did occur. Thus, had any one of the steps of the events involved in the flight of Sylvia's ball before the collision (D) been altered or not occurred, the shattering would still have occurred thanks to Bruno's throw.

*The Humean counterfactual analysis fails*: Recall the details of the Humean counterfactual analysis; where C and E are distinct actual events, then:

- C causes E iff E counterfactually depends on C.

The analysis fails to analyse Sylvia's throw as a cause of the shattering (E), as follows:

- The shattering (E) does not counterfactually depend on Sylvia's throw (C). Or in other words, if Sylvia's throw had not occurred (C), then the shattering still would have occurred (E), because Bruno's ball would have caused an exact same shattering (E).

We note that the Humean analysis analyses Bruno's throw as a cause; if Bruno's throw (A) had not occurred, then there would have been no shattering (E), given there would

then have been no collision (D), Sylvia's ball would have landed near her intended target (something that wasn't the bottle), and the bottle would not have been shattered.

We can see that the strategy of appealing to 'fine-grained-ness' of the effect won't help us here as regards the Humean analysis, given that the shattering which Bruno's ball would havecaused, had Sylvia not thrown, would have been exactly the same as the shattering Sylvia's throw actually caused (we supposed this in the example above). How do Lewis' strategies fare?

*The counterfactual-chains analysis fails*: We have supposed that any alteration to Sylvia's throw would have made the collision impossible, and that if Sylvia had not thrown, the shattering would still have occurred in the exact same fragile fashion as it actually did. Might there be a way to appeal to chains of dependence to fix the problem? Recall the details of Lewis' counterfactual chains analysis:

- C causes E iff there is a counterfactual chain of dependence leading from E to C.
- There is a counterfactual chain of dependence from E to C iff there is a set of actual events E, D, $D_1$ etc… C, such that E counterfactually depends on D, D on $D_1$ etc…. C.

Strevens observes that no appeal to chains will help us in this example. He observes that no event to do with Sylvia's ball's passage after the collision counterfactually depends on any event to do with Sylvia's ball flying toward the collision or the collision itself. The

observation is correct; had any event of the trajectory of Sylvia's ball before or including the collision not occurred, an intrinsically identical set of events would have occurred after the collision up to the shattering, courtesy of Bruno's throw. Thus the counterfactual chains analysis fails to analyse Sylvia's throw as a cause.

*The dependence fixing analysis succeeds*: Recall the details of the dependence fixing analysis: Where C, E, and F are distinct actual events, then:

- C causes E iff there is some event F such that if F had still occurred and C not, then E would not have occurred.

The analysis succeeds in analysing our example, as follows:

- There is some event F (i.e. the events which constitute the trajectory of Bruno's ball), such that if F had still occurred and Sylvia's throw (C) not, then the shattering (E) would not have occurred.

Admittedly, the dependence fixing analysis achieves its goal in a strange way. That is, the counterfactual worlds we appeal to in order to achieve our analysis are worlds in which Bruno's ball seems to take a very unnatural trajectory, Bruno's ball takes a right turn in mid-air and without apparent cause (given Sylvia's ball is no longer present)! This shouldn't worry us too much, given Lewis' similarity relation allows for 'local violations' of laws in such a way as not to damage the overall layout of laws in that

138

world; and we can confirm this by direct inspection of Lewis' similarity relation (it is only our second last priority to maintain local non violations of laws, but our first priority to maintain global non violations of laws).

*The dependence fixing analysis fails with appeal to a revised example*: the dependence-fixing analysis doesn't seem to analyse some strengthened versions of the Sylvia and Bruno example. To see this, consider the following revised example:

### 3.3.2 Strengthened middle pre-emption

**Example (21)**:

SYLVIA AND BRUNO#2 (also due to Strevens [2003]): Suppose the details of SLYVIA AND BRUNO#1 (example 20, p. 129) again. Except this time, suppose Sylvia was in fact aiming for a wall, in the hope that her ball would rebound off that wall, and off another wall standing at a perpendicular angle to the first wall, and then smash the bottle (so suppose Sylvia was trying to be more skilful in her destruction of the bottle than Bruno - and that this plan would have succeeded, had such a process remained uninterrupted). Suppose Bruno was directly aiming for the bottle, but that Sylvia and Bruno's balls collide again, rendering it true that Sylvia's ball shatters the bottle (E) without hitting any of the walls, but also true that Bruno's ball reaches the shattered bottle half a second later than Sylvia's ball does. Again, both Sylvia's and Bruno's throw were causes of the

shattering - Sylvia's thrown ball shattered the bottle, and Bruno's throw caused the collision that enabled Sylvia's thrown ball to do so.

A topographical diagram will help illustrate this revised example. The black line represents the trajectory of Sylvia's ball, the dotted line represents the trajectory of Bruno's ball. Had Bruno not thrown his ball, Sylvia's ball would have taken the trajectory of Bruno's ball after the collision, and caused a similar but delayed shattering.

Sylvia's ball            ◯  Glass bottle

Bruno's ball
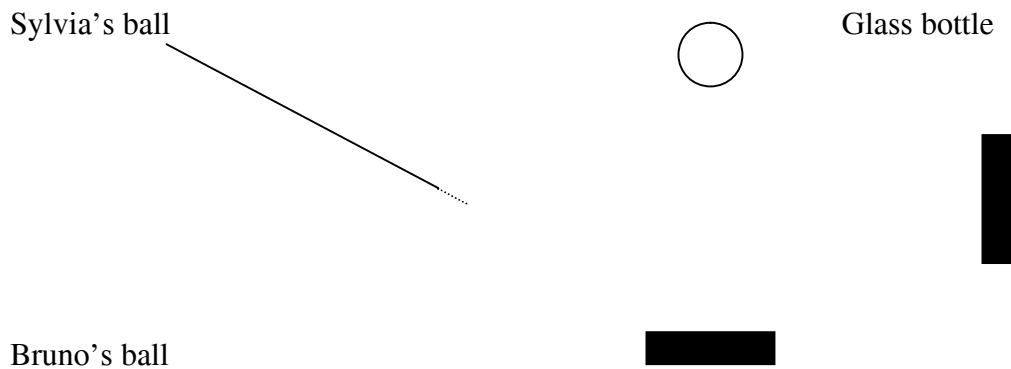
What's the problem for the holding fixed analysis? The problem is that there are no events that we can 'hold fixed' in the appropriate way in order to yield the correct result that Sylvia's throw (C) was the cause of the shattering (E). For the first example, our analysis succeeded because we could hold fixed the events that constituted the trajectory of Bruno's ball to some time after the collision, and see that the closest non-Sylvia throwing worlds where these events still occurred were non-shattering worlds. In this case, holding fixed the same events renders it true that Bruno's ball causes a similar shattering in a similar way by an alternative route; on the other hand, not holding those

events fixed renders it true that if Sylvia hadn't thrown then Bruno's ball would have caused the exact same shattering.

Can an appeal to fine-grained events help here as a possible modification for the dependence fixing analysis? Yes. Taking E in the causal analysis as a fine-grained event means we can appeal to the fact that had we held fixed the events which constituted Bruno's ball's actual trajectory, then the closest non-Sylvia throwing worlds are worlds in which the fragile version of the shattering would not have occurred. For starters, the shattering Bruno's ball would have caused would have been slightly later (half a second), and would have occurred in a different manner (from the opposite side of the bottle). Thus, fine-graining the effect in the holding fixed analysis has proved a success for cases of late and middle pre-emption.

*The influence analysis fails in both examples*: Recall the details of the influence analysis; where C and E are distinct actual events:

- C causes E iff C influences E.
- C influences E iff there is a substantial range of $C_1 \ldots C_n$ of different not-too-distant alterations of C (including the actual alteration of C) and a range of $E_1 \ldots E_n$ of alterations of E, at least some of which differ, such that if $C_1$ had occurred, $E_1$ would have occurred, if $C_n$ had occurred, $E_n$ would have occurred.[92]

---

[92] Lewis, [2003], p. 91

The influence analysis fails to analyse Sylvia's throw in either example: there is no close range of not-too-distant alterations to Sylvia's ball which counterfactually imply a range of alterations to the shattering. This is because if Sylvia's throw had been altered in any close fashion, Sylvia's ball would have missed Bruno's ball, and so Bruno's ball would have caused an exact same fragile version of the shattering Sylvia's ball actually caused (or we may suppose suitably contrived circumstances to safely assume this). How does the influence chains analysis fare?

*The influence chains analysis fails*: We might try and appeal to chains of influence in order to identify Sylvia's throw (C) as the cause of the shattering (E). We might try to do this by proposing that although Sylvia's throw has no influence on the shattering of the bottle itself, it has an influence on an intermediary event D (the collision), which itself has an influence on the shattering. If this were true, then we could identify Sylvia's throw as the ancestral influence on the shattering, and thereby a cause of the shattering.

This proposal fails. Strevens observes that the example can be set up such that the balls are so small and fast[93] that any alteration in the collision would result in the balls not colliding at all, to the effect that the collision does not influence the shattering, because Bruno's ball would then take Sylvia's trajectory at time t and cause the shattering in the same time and manner as Sylvia's actually did. Consequently, Sylvia's throw is not linked to the shattering as the ancestral influence by appeal to D alone.

---

[93] He suggests that perhaps they are fired from guns.

The problem Strevens identifies here is that any chain of events that involves alterations of D would result in no collision. As Strevens observes, the shattering of the bottle depends on the events (of the ball states) after the collision, and the collision depends on events before the collision, but no events after the collision depend on the collision or any event before the collision, since had there been no collision, an intrinsically identical chain of post collision events would have led up to the shattering of the bottle,[94] as Bruno's ball, intrinsically identical to Sylvia's, would have assumed the other ball's trajectory after t.

Choi [2005], in a reply to Strevens, argues that there are some factors that Strevens does not account for that make altered versions of the collision possible (for instance, making the collision have a "neutral electromagnetic charge" or making Sylvia's ball "blue", where the actual collision involves silver steel balls with a non neutral electromagnetic charge). In which case we might identify a transitive chain of influence as follows:

- An alteration of the shattering E counterfactually depends on an alteration Z of an event just before the shattering (say an altered shattering is counterfactually dependent on the ball being "rectangular" just before the shattering).

- An alteration of Z counterfactually depends on the collision D (say the blue ball in altered event Z is counterfactually dependent on the ball being "blue" in altered event D)

---

[94] Strevens, [2003], p. 402

- An alteration of D is counterfactually dependent on an event just before the collision Y (say an altered collision with "neutral electromagnetic charge" counterfactually depends on the ball having "neutral electromagnetic charge").

Choi's response illustrates how Strevens' example, as presented, is fundamentally flawed if it is meant to be a counter-example to the necessity of the influence chains analysis. The flaw is that Strevens allows us to appeal to a more detailed chain of stepwise causal influence that he had not anticipated. Although Choi's observations seem to be correct about the example Strevens gives, I think Choi's objection misses its target:

We suppose that a point of Strevens' case is that it is possible to set up the example in which no events after the collision depend or are influenced by the collision or any event before the collision in the range of close possible worlds. Following this thought, we can easily make additional revisions to the example in which any alterations to the ball's colour, mass, electric charge etc. (these are three of Choi's suggestions which he spends some length discussing) would have made the collision impossible; perhaps by appeal to factors like heat absorption (regarding the colour), gravitational factors (regarding the mass), and strong magnetic fields (regarding the electric charge). Note, we don't need to make these revisions remotely realistic, indeed we can make them as unrealistic as we like (we can appeal to "laws of magic" like example (4) does); merely the intuition that a causal process occurs between C and E must remain in our example. We can state such an example here:

**Example (22):** SYLVIA AND BRUNO#3: Suppose the details of SYLVIA AND BRUNO#1 (example 20, p.129) again; this time, suppose that the affair is watched closely by robots who would have i) destroyed Sylvia's ball had her throw or her ball been altered in any fashion whatsoever on its passage up to the shattering, and ii) permitted an identical shattering to the actual shattering via Bruno's ball.

On this strengthened example, it is obvious that neither the influence nor the influence chains analysis succeeds in analysing Sylvia's throw as a cause. Is there a way to revise the influence chains analysis to accommodate the view that Sylvia's throw was indeed a cause of the shattering? Strevens considers one such revision, which we shall term the 'maximal influence chains analysis'; where C and E are distinct actual events, then:

- C causes E iff there is a chain of maximal influence from E to C

- There is a chain of maximal influence from E to C iff C maximally influences D, D maximally influences D1 etc… E

- C maximally influences E iff C has a relatively large influence on E (relative, that is, to the influence of the other causal influences on E).

This revised analysis does us no good here - the problem remains that there is no chain of influence simpliciter from C to E in example (22) - and thus no chain of relatively large influences. Furthermore, the maximal influence chains analysis suffers from the same problem as the other influence analyses; that of being insufficient for causation. Here is an example, which demonstrates this:

**Example (23):** ELEPHANT HUNTER (a variation of Strevens' example[95]) Suppose that Sylvia and Bruno's inebriated mother is hunting pink elephants nearby with a very high powered blunderbuss, and suppose she fires her gun and its shot lands a small distance away from the bottlejar (suppose she thought the bottle was a pink elephant, and she only just missed her 'target'), just as Sylvia and Bruno throw their balls at the bottle. Also suppose that if her shot had hit the bottle (which as it is stands is the closest alteration to her actual shot, given she only missed by the smallest of incidences), it would have made it explode in a tremendous fashion, making far more of an alteration to the event of the shattering than either of Sylvia or Bruno's throws.

As we can see, the influence, influence chains, influence fixing and maximal influence chains will all deliver the incorrect result that Sylvia's mother's shot was a cause of the shattering. Thus we achieve our general result that no version of the influence analysis met so far is necessary or sufficient for causation in cases of middle pre-emption. However, the influence fixing analysis has better results for examples of middle pre-emption. We shall not visit how here, given it is obvious it should succeed in example (20), given we can appeal to close possible worlds in which Bruno's ball's trajectory is 'held fixed' and then test for influence, but should fail for similar reasons as the dependence fixing analysis does in example (21). We shall now visit how the revised quasi-dependence and revised would-be dependence analyses succeed in our middle pre-emption examples:

---

[95] Strevens, [2003], p. 407

*The revised quasi-dependence analysis succeeds*: Recall the details of the revised quasi-dependence analysis:

- C causes E iff E quasi-depends on C

- E quasi-depends on C iff there is an isonomic duplicate of a process from E to C, in which the duplicate of E counterfactually depends on the duplicate of C.

- X is an isonomic duplicate of Y iff X and Y have exactly the same events, and the same governing laws.

- Two duplicate sets of events X and Y have the same governing laws iff it's the case that events A and B are in X and A nomologically determines B iff events A' and B' are in Y and A' nomologically determines B'.

The revised quasi-dependence analysis succeeds in analysing Sylvia's throw as a cause of the shattering. There are many different isonomic duplicates (of the sort analysed above) to which we can appeal in our analysis to make the conclusion that Sylvia's throw caused the shattering true. First, we might take a world which includes an isonomic duplicate of the entire states of affairs of Bruno and Sylvia, where Billy (Bruno's younger brother) would have caught Bruno's ball before it had any chance to get near the glass bottle. Second, we might appeal to an isonomic duplicate of the process running up from Sylvia's throw to the shattering, but in which Sylvia's ball collides with an immobile wall instead of Bruno's ball. In either duplicate, we can analyse the shattering as counterfactually dependent on Sylvia's throw.

Note that all that is required between isonomic duplicates in our revised (non-Lewisian) notion is the same relevant pattern of nomological determinacy between each duplicate - we need not preserve laws which determine events outside the duplicate from inside the duplicate. As such, we avoid revenge versions of the 'problem of back-up determiners' fitted to cases of middle pre-emption; a problem that would otherwise confront Lewis' original quasi-dependence analysis.

*The revised would-be dependence analysis succeeds*: Recall the details of the revised would-be dependence analysis:

- C causes E iff there is a chain of counterfactual dependence linking C to E, or there would be such a chain had it not been for some pure dependence preventer, and the would-be chain still occurred

- An event P purely prevents dependence of E on C iff P is a dependence preventer on a chain of events from E to C, and P prevents no event in a chain of events that would form a counterfactual chain linking E to C had P not occurred.

- P is a dependence preventer of E on C iff had P not occurred and the C-E chain of events still had, then E would counterfactually depend on C.

The revised would-be dependence analysis succeeds in analysing Sylvia's throw as a cause of the shattering, as follows:

- There would be a chain of counterfactual dependence linking the shattering (E), to Sylvia's throw (C), had it not been for some pure dependence preventer (namely, Bruno's throw), and the would-be chain (from Sylvia's throw to the shattering) still occurred

- Bruno's throw (A) is a pure dependence preventer of the shattering (E) on Sylvia's throw. Firstly, we note it is a dependence preventer given that had the C-E chain of events still occurred and Bruno's throw not occurred, then the shattering would have depended on Sylvia's throw; and secondly, it is a *pure* dependence preventer given it prevents no event in a C-E chain from actually occurring.

Note that Collins' original analysis would have failed to analyse Sylvia's throw as a cause of the shattering, namely because Collins excludes the revision in the analysis stated in the last conjunct of the first bullet point; *and the would-be chain still occurred.* In our example, there would be no such chain had it not been for some pure dependence preventer simpliciter. This is because, had it not been for the pure dependence preventer of Bruno's throw, then the collision would not have occurred, and there would not be a chain of counterfactual dependence linking the shattering to Sylvia's throw (given there would have been no shattering, and no collision to provide for this!).

*Some observations*: We catalogue the projected and demonstrated results for each analysis as they deal with our cases of middle pre-emption. We visited two main examples (20) and (21):

- The Humean counterfactual analysis fails in all cases

- The holding fixed and holding fixed influence analyses succeed in all cases but 21

- The fine-grained holding fixed analysis succeeds in all cases

- The counterfactual chains analysis fails in all cases

- The influence, maximal influence, and influence chains analyses fail in all cases

- The revised quasi-dependence analysis succeeds in all cases

- The revised would-be dependence analysis succeeds in all cases

Once again, strategies of 'holding fixed' and 'getting rid of' come up trumps. However, to presume success for such strategies, as we shall now see, is premature, for where they do not come up trumps is in cases of trumping pre-emption; one of the most vicious examples of pre-emption yet to be seen.
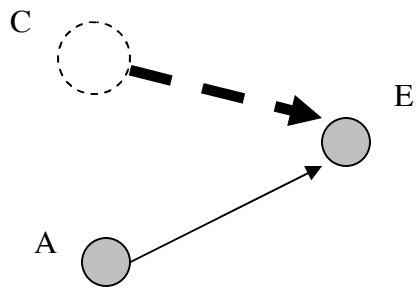
## 3.4 Trumping pre-emption

*Intuitive explanation*: Cases of trumping pre-emption are cases of pre-emption in which the process running from the pre-empted back-up cause to the effect is intrinsically just like a causal process, except extrinsic factors determine the process as causally inefficacious. Cases of trumping are due to Schaffer (2000).

*Diagrammatic re-construction*: (big grey circles and big black arrow represent a 'trumping' causal process, dashed thin arrows represent 'trumped' causal processes,

empty big circles and dashed big arrows represent trumping causal processes that don't occur).

C

E

A
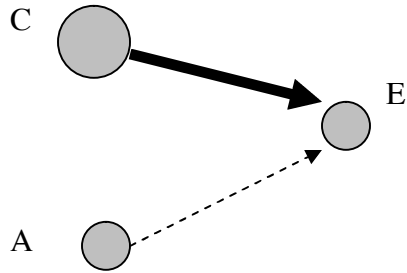
The actual world, C causes E, and C 'trumps' A's process from A to E

C

E

A

The closest worlds in which C doesn't occur, A (no longer 'trumped' by C) causes E

## 3.4.1 Basic trumping pre-emption

**Example (24):**

*First example* (due to Schaffer[96]): Suppose a major shouts "charge" to the corporal (C), and a sergeant simultaneously shouts "charge" to the corporal as well (A). Suppose the corporal charges (E). The corporal's charging would have occurred as a result of the major's orders alone, or the sergeant's, but when they both occur together, the major's shouting is the exclusive factor which causes the charging (it 'trumps' the sergeant's orders), given a corporal will always obey the orders of the highest ranking soldier with a rank higher than him.

**Example (25):**

*Second example* (also due to Schaffer[97]): Schaffer observes we can also give 'trumping' examples in which the actual cause precedes the trumped cause (as an interesting variation of cases of trumping). Suppose it is a law of magic that the first spell cast on any given day matches the following enchantment at midnight. Merlin casts the first spell that day to turn the prince into a frog (C), Morgana casts the same spell slightly later (A), and at midnight, the prince turns into a frog (E). Schaffer observes that clearly Merlin's spell, and not Morgana's, is the exclusive factor which causes the enchantment (it 'trumps' Morgana's spell), because the laws provide for the determination of the enchantment by Merlin's spell and not by Morgana's.

**Example (26):**

---

[96] Schaffer, p. 59, [2000]

[97] Schaffer, p. 67, [2000]

*Third example* (also due to Schaffer[98]): Suppose a certain type of particle takes a certain type of curved trajectory (E) when it is subject to a certain type of field (called "black", "grey", or "white") with a certain sort of magnitude. Suppose also that whenever such a particle is subject to multiple fields scattered around the particle, it will take a trajectory as if only subjected to the magnitude of the "darkest" field. Suppose that such a particle is subject to two fields next to one another (one black (C), the other white (A)), such that the particle takes the trajectory (E) as determined by the black field (C), but would have taken the same trajectory had the black field not been there, given the presence of the white field (A).

First of all, what are we to make of cases of trumping? Hall and Paul are two philosophers who reject Schaffer's view that only C is the cause of E in such cases. They write:

> "Although they have attracted a small flurry of interest in the recent literature, cases of trumping turn out on inspection to be nothing more than either cases of symmetric over-determination in disguise, or cases of late pre-emption in disguise; either way, they have nothing new to teach us."[99]

---

[98] Schaffer, p. 67, [2000]

[99] Hall and Paul, p.3, [2006]

153

No clue is given as to why they think this. We might suppose that they think that if they are cases of late pre-emption in disguise, then we can identity some intrinsic properties in the process running from the pre-empted back-up cause that fail to 'get in touch' with the effect (as this is characteristic of cases of late pre-emption). This may be the case with the sergeant-major example - perhaps we can indeed identify some physical process from the sergeant's order that doesn't 'get in touch' with the corporal's motivating thoughts to charge; but not enough is known about the physics of such cases to make a determinate answer. In any case, the Merlin-Morgana case is clearer on this matter - there seems no process that is physically 'cut short' from 'reaching' the effect as far as the pre-empted cause (Morgana's spell) is considered, and so contra Hall and Paul, at least this example isn't simply a case of late pre-emption.

What about the alternative suggestion, that such cases could be analysed as cases of over-determination? We shall meet these cases later, but for now all we need to know is that if Schaffer's cases are actually cases of over-determination, then both C and A are to be analysed as causes of E (contra Schaffer's conclusion that only A is not a cause of E). This suggestion has some weight to it; the conclusion that A is also a cause of E is certainly not as implausible as the suggestion that the pre-empted back-up cause is actually a cause of the effect in the other cases of pre-emption; for starters it seems as if the 'process' from Morgana's spell to the enchantment is *intrinsically just like* causally efficacious counterparts to her spell. Following this, and as far as brute intuitions are concerned, I would say it would be fortunate if an analysis analysed A as a non cause of E, but not a disaster if it didn't.

*The Humean counterfactual, dependence fixing, and counterfactual chains analyses fail*:
It's easy to demonstrate how these three analyses fail, so we shall deal with them in one go; we shall demonstrate this by appeal to the Merlin/Morgana example:

First, the Humean counterfactual analysis fails because the enchantment (E) does not counterfactually depend on Merlin's spell (C). This is because had Merlin's spell not occurred, then the enchantment still would have, because Morgana's spell would have occurred to cause the enchantment.

Secondly, the counterfactual chains analysis fails because there is no chain of events running from the enchantment (E) to Merlin's spell (C) which constitutes a chain of dependence. This is because Merlin's spell causes the enchantment directly by appeal to 'action at a spatio-temporal distance', without appeal to any intermediary events, and thus there is no event the enchantment depends on which is linked to a chain of dependence ending up with Merlin's spell.

Thirdly, the dependence fixing analysis fails because there is no event that can be 'held fixed' such that if Merlin's spell hadn't occurred, the enchantment wouldn't have; in any such variation of 'holding fixed', Morgana's spell would still have caused the enchantment.

*The influence and influence-chains analyses seem to succeed:*[100] We note that the influence and influence chains analyses seem to analyse Merlin's spell (C) as a cause of the enchantment (E), and Morgana's spell as a non cause. To see this in depth, recall the details of the influence analysis; where C and E are distinct actual events, then:

- C causes E iff C influences E, and C and E are distinct actual events.

- C influences E iff there is a substantial range of $C_1 \ldots C_n$ of different not-too-distant alterations of C (including the actual alteration of C) and a range of $E_1 \ldots E_n$ of alterations of E, at least some of which differ, such that if $C_1$ had occurred, $E_1$ would have occurred, if $C_n$ had occurred, $E_n$ would have occurred.[101]

We note:

1. Merlin's spell (C) influences the enchantment of the prince turning into a frog (E). Or, in other words, there is a substantial range of not-too-distant alterations of Merlin's 'prince to frog' spell, which counterfactually implies a range of alterations to the enchantment. As follows: Had Merlin cast the spell 'prince to newt' ($C_1$), then the prince would have turned into a newt ($E_1$), had Merlin cast the spell 'prince to worm' ($C_2$), then the prince would have turned into a worm

---

[100] Demonstrating that the influence analysis succeeds in this case indirectly demonstrates that the influence chains analysis succeeds; given it is true that if an event influences another, it influences it via a chain of events.

[101] Lewis, [2003], p. 91

($E_2$), had Merlin cast the spell 'prince to kangaroo' ($C_3$), then the prince would have turned into a kangaroo ($E_3$)… etc.

We note here that Morgana's spell does not influence the enchantment, given that the not-too-distant alterations of her spell (we can consider the range of Merlin's spells above if we like) do not counterfactually imply a range of alterations to the enchantment. This is simply because in the range of worlds in which Morgana's spell is not too distantly altered, the law of nature 'the first spell cast that day matches the enchantment at midnight' still holds, and thus Merlin's 'prince to frog' spell would still have influenced the prince to turn into a frog.

We should be sufficiently accustomed to variations of 'modally fragile' pre-emption to see how the influence analysis can be made to go wrong. For that, we consider an example of Rosen's:

### 3.4.2 Strengthened trumping pre-emption

**Example (27):** *The influence chains analysis fails upon a revised example* (due to Rosen[102]): Suppose a similar example to Schaffer's example about Merlin and Morgana, except for the following; suppose that the laws of magic distinguish the magical powers of a wizard from those of a witch; also suppose that Merlin can only cast one type of efficacious spell (that of turning a prince into a frog), and can only cast one spell at one

---

[102] See Collins, [2000], p. 104

time of day (at noon). Thus, Merlin can only cause the effect at one unique time and manner in the range of not-too-distant possible worlds. Suppose also that the governing law of magic is no longer 'the first spell cast that day matches the enchantment at midnight', but a similar law - 'the first prince-to-frog spell that day matches the enchantment at midnight'.

It is easy to see how the influence analysis fails in this revised example: there is *no* substantial range of alterations to Merlin's spell (C), which counterfactually implies a range of alterations to the enchantment (E). Had Merlin's spell been altered in any not-too-distant way (we may consider the range of alterations we gave for our last example here if we like), then the prince would still have turned into a frog - Morgana's spell would have been the first prince-to-frog spell that day, and thus would have determined the prince-to-frog enchantment, given the laws.

Thus the influence analysis is demonstrated as insufficient for causation. We note that Rosen's example is very similar to the sort of 'modally fragile pre-emption' I gave in cases of early and late pre-emption (recall that cases of modally fragile pre-emption endeavoured to show how something may cause an effect without influencing it). Rosen also revises his example to give an example of the sort of 'over-generative pre-emption' examples I gave earlier. Here is his extended example:

**Example (28):** A variation of an example from Rosen[103]: Suppose the details of example (26), except this time, along with supposing that Merlin can only cast one type of spell at one particular time and manner, suppose the governing law of nature to be the law 'the first prince-to-frog spell that day matches the enchantment that night, unless there are spells which say otherwise'. Again, Merlin and Morgana both cast a prince-to-frog spell, and so Merlin's spell 'trumps' Morgana's spell.

It is easy to see how the influence analysis is unnecessary for causation in this case; not only is it true that Merlin's spell does not influence the enchantment (for the same reason as it didn't in the last example), it is now also true that Morgana's spell *does* influence the enchantment - not-too-distant alterations of Morgana's spell counterfactually imply a range of alterations to the enchantment, given the laws. We are now left with the wholly unsatisfactory conclusion that Morgana's spell, and not Merlin's, caused the enchantment. Thus the influence analysis is neither necessary nor sufficient for causation in trumping examples. How do our quasi-dependence and would-be dependence analyses fare?

*The revised quasi-dependence analyses analyse cases of trumping as cases of over-determination*: Recall the details of the quasi-dependence analysis:

- C causes E iff E quasi-depends on C

---

[103] Collins, [2000], p. 104

- E quasi-depends on C iff there is an isonomic duplicate of a process from E to C, in which the duplicate of E counterfactually depends on the duplicate of C.

The quasi-dependence analysis analyses cases of trumping as cases of over-determination (namely, it analyses both Merlin's and Morgana's spells as causes of the enchantment). As follows:

1. The enchantment (E) quasi-depends on Merlin's spell (C)
2. There is an isonomic duplicate of the process running to the enchantment (E) from Merlin's spell (C) (namely, in a world where Morgana's spell (A) does not feature), in which the enchantment (E) counterfactually depends on Merlin's spell (C).

3. The enchantment (E) quasi-depends on Morgana's spell (A)
4. There is an isonomic duplicate of the process running to the enchantment (E) from Morgana's spell (A) (namely, in a world where Merlin's spell (C) does not feature), in which the enchantment (E) counterfactually depends on Morgana's spell (A).

Thus, both Merlin and Morgana's spells are analysed as causes of the enchantment. We note here that we have been quite charitable as to what constitutes a process. Namely, we have said that there is a process from Merlin's and Morgana's spells to the enchantments (albeit with the supposition that only the process from Merlin's spell was causally

efficacious), but this is far from clear. Traditionally at least, processes seem to be chains of contiguous events which don't admit of causation at a spatio-temporal distance. However, given that quasi-dependence fails to provide us with truth conditions for what a process actually is, we are at a loss in deciding whether we should be saying there are such 'magic spell' processes of the sort required for the quasi-dependence analysis to analyse Merlin's spell as a cause of the enchantment in this case.

Before drawing some general observations about these conclusions, we shall first examine how the would-be dependence analysis fares.

*The revised would-be dependence analysis analyses cases of trumping as cases of over-determination*: Recall the details of the would-be dependence analysis:

- C causes E iff there is a chain of counterfactual dependence linking C to E, or there would be such a chain had it not been for some pure dependence preventer, and the would-be chain still occurred

- An event P purely prevents dependence of E on C iff P is a dependence preventer on a chain of events from E to C, and P prevents no event in a chain of events that would form a counterfactual chain linking E to C had P not occurred.

- P is a dependence preventer of E on C iff had P not occurred and the C-E chain of events still had, then E would counterfactually depend on C.

The would-be dependence analysis analyses cases of trumping as cases of over-determination (namely, it analyses both Merlin's and Morgana's spells as causes of the enchantment). As follows:

1. The enchantment (E) would counterfactually depend on Merlin's spell (C), had it not been for Morgana's spell (A), and the would-be chain (Merlin's spell and the enchantment) still occurred

2. Morgana's spell (A) purely prevents dependence of the enchantment (E) on Merlin's spell (C), firstly because it prevents this dependence (had her spell not occurred and Merlin's spell and the enchantment still occurred, then the enchantment would have counterfactually depended on Merlin's spell); and secondly because it prevents no event in a chain of events that would form a counterfactual chain linking the enchantment to Merlin's spell.

A similar analysis also holds for Morgana's spell (we can easily see this by running through the above and simply swapping around the names of Merlin and Morgana).

Are the results that the quasi- and would-be dependence analyses produce correct? We said earlier that it wouldn't be a complete disaster if cases of trumping were analysed as cases of over-determination, given the very strong intrinsic resemblance of Morgana's spell-processes to cases in which we are sure that her process was surely causally efficacious. The would-be dependence analysis achieves the same result for obvious reasons.

*Some observations*: We catalogue the projected and demonstrated results for each analysis as they deal with our cases of trumping. We visited three of Schaffer's examples, and two revised examples courtesy of Rosen. Our results are as follows:

- The Humean counterfactual, dependence-fixing, and counterfactual chains analyses fail in all cases

- The influence, maximal influence, influence chains, and influence fixing analyses succeed in Schaffer's cases

- The influence, maximal influence, influence chains, and influence fixing analyses fail in Rosen's cases

- The revised quasi-dependence and would-be dependence analyses generate the permissible conclusions that trumping cases are cases of over-determination.
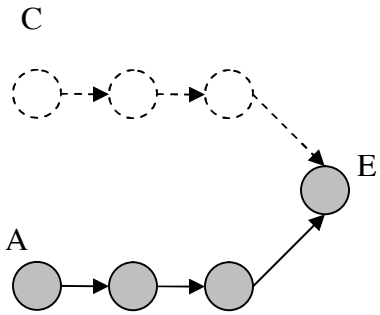
## 3.5 Double Pre-emption

*Intuitive explanation*: In cases of symmetrical double pre-emption (otherwise known as cases of symmetrical over-determination), an effect has at least two causes each of which is causally sufficient for the effect. (See McDermott[104]):
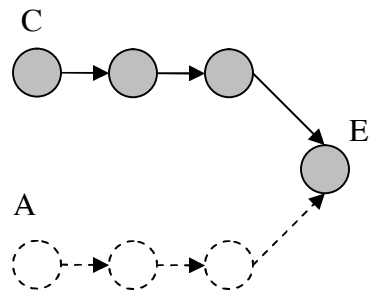
---

[104] McDermott, [1995], p. 528
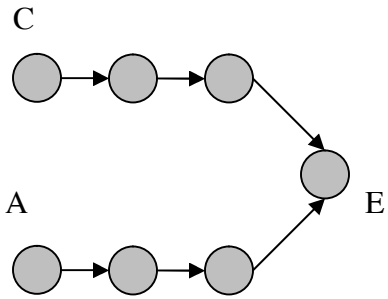
The actual world, C causes E via a chain of events, as does A.



The closest possible worlds in which C does not occur, A still causes E via a chain of events.

The closest possible worlds in which A does not occur; C still causes E via a chain of events.

### 3.5.1 Basic double pre-emption

**Example (29):** BILLY AND SUZY#1': Suppose a familiar example of Billy and Suzy throwing rocks at a window. Suppose Billy throws a rock at the window (A), and at the same time Suzy throws a rock at the window (C) as well. The window shatters (E), caused by both throws, each of which was sufficient for the window shattering (E).

We should be sufficiently equipped to see how the basic Humean, dependence fixing, and counterfactual chains analyses fail to analyse either A or C each as an independent cause of E. E does not counterfactually depend on C or A alone, nor would it with some events held fixed, and there isn't a 'counterfactual chain' of dependence from the effect to either cause.

We also suppose that the influence, influence chains, and influence fixing analyses succeed in these examples, all things being equal. This is because there seems to be a pattern of not-too-distant alterations to Billy's/Suzy's throw, that counterfactually implies a range of alterations to the shattering, with regard to how the shattering occurs. The fine-grained Humean counterfactual analysis also seems to succeed, given that the fragile version of the shattering would not have occurred had either one of the throws not occurred.

However, here is a case of over-determination that is a counter example to all of the aforementioned analyses.

### 3.5.1 Strengthened double pre-emption

**Example (30):** LIGHT BULB (my own example): Suppose that for the street lights to turn on at precisely midnight, either a button can be pressed in the town hall at exactly 6pm the previous evening, or a button must be pressed at the exact same time at the power station. Suppose that pressing either button ensures the lights turn on at precisely midnight in a specific manner. Suppose the button is pressed at town hall at 6pm (C), and as it happened, the button is pressed at the same time at the power station (A), such that the street lights turn on at midnight (E). We suppose in this example that both pressings of the button cause, and are causally sufficient for, the fragile version of the lights coming on at midnight.

*The fine-grained Humean counterfactual analysis fails*: Had the button pressing at the town hall (C) not occurred, the fragile version of the lights turning on at midnight (E) still would have occurred, because the button pressing at the power station (A) would have sufficed for the lights turning on.

*All the variations of the influence analysis fail*: We shall only demonstrate that the influence analysis fails for this example here, given it should now be obvious how its variations will fail. Recall the details of the influence analysis:

- C causes E iff C influences E, and C and E are distinct actual events.
- C influences E iff there is a substantial range of C1… Cn of different not-too-distant alterations of C (including the actual alteration of C) and a range of E1… En of alterations of E, at least some of which differ, such that if C1 had occurred, E1 would have occurred, if Cn had occurred, En would have occurred.[105]

The influence analysis fails to analyse example (30) correctly; it neither analyses the town hall pressing (C) nor the power station button pressing (A) as each an independent cause of the lights turning on at midnight (E). Take either button pressing; it is not the case that not-too distant-alterations of one of the pressings (pressing earlier, later, harder, softer.. etc.) counterfactually imply a range of alterations to the effect. This is because had one of the pressings been altered in any not-too-distant fashion, the other button pressing would have sufficed for a fragile version of the actual event of the street lights coming on at midnight (E).

It is also obvious that no appeal to the fragile version of the lights turning on at midnight (E) will render that event counterfactually dependent on either one of the two button

---

[105] Lewis, [2003], p. 91

pressings; had only one of the button pressings not occurred, then the same fragile version of the lights turning on at midnight would have occurred, and thus the fine-grained Humean counterfactual analysis fails to analyse each of the button pressings as a cause of the light turning on at midnight (E).

*The quasi-dependence and would-be dependence analyses succeed*: Better results are had with the revised quasi-dependence analysis and the revised would-be dependence analysis. In the first instance, for each of the button pressings, there is an isonomic duplicate of that button pressing and its process leading up to the street lights turning on, in which the lights turning on counterfactually depend on that pressing. In the second instance, for each of the button pressings, the street lights turning on (E) would have counterfactually depended on that pressing, had it not been for the pure dependence preventer of the other button pressing, and the would-be chain from that pressing to the lights turning on still occurred. Thus, both the revised quasi-dependence and the revised would-be dependence analyses succeed in analysing both pressings each as a cause of the street lights turning on (E).

*Some observations*: We catalogue the projected and demonstrated results for each analysis as they deal with our cases of symmetrical over-determination. We visited two main examples; our results are as follows:

- The Humean counterfactual, holding fixed, and counterfactual chains analyses fail in all cases

168

- The influence, maximal influence, influence chains, and influence fixing analyses fail in example (30), but succeed in (29)

- The revised quasi-dependence and revised would-be dependence analyses succeed in all examples.

So far we have met with our most menacing case of pre-emption - that of 'double pre-emption', otherwise known as 'symmetrical over-determination'. As we can see, 'getting rid of' strategies prosper over 'holding fixed' strategies with respect to such cases, where they were on a more or less equal footing with respect to cases of early and late pre-emption. Is there an easy fix for the counterfactual strategy with respect to symmetrical over-determination cases? Bunzl once made certain observations about such cases in order to weaken their grip over the counterfactual strategy. It is worth investigating these observations in full, as they will lead to some important insights into the counterfactual strategy in general.

## 3.6 Bunzl pre-emption

In 'Causal Over-determination',[106] Bunzl observes that in typical cases of symmetrical over-determination, we can often identify a 'Bunzl event', an analysis of which is provided here:

- An event B is a Bunzl event iff

---

[106] Journal of Philosophy 76 [1979].

There are two events, C1 and C2, which causally over-determine an event E, such that C1 and C2 cause B (without C1 and C2 together causally over-determining B), and B causes E, (without B being a causal over-determiner of E).

Bunzl's observation is that in cases of over-determination, such 'Bunzl events' can be identified as causal intermediaries between the over-determined effect and the over-determining causes. The observation is interesting, as it can be used to show how a lot of cases of symmetrical over-determination 'hide' cases of early and late pre-emption, in such a way that allows counterfactual analyses of causation that succeed in cases of early and late pre-emption to analyse them away. To see the merits of this observation, we visit two cases of 'Bunzl pre-emption'. The first is a case of 'Bunzl late pre-emption' (or, as Lewis calls it, a case of 'self-pre-emption'), and the second is a case of 'Bunzl early pre-emption'. We shall then make some observations about how these cases reflect on the counterfactual analyses of causation we have been visiting.
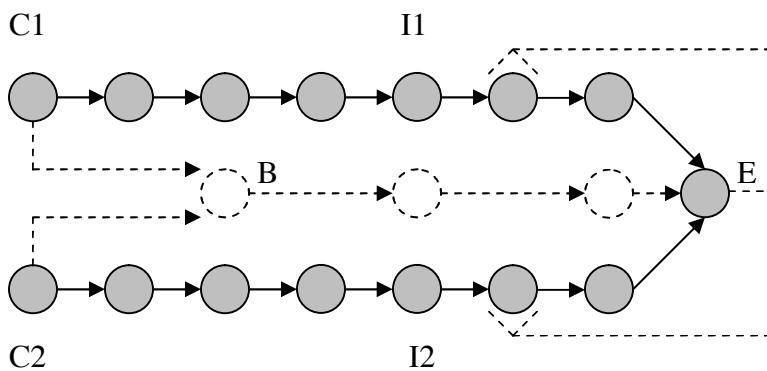
### 3.6.1 Bunzl late pre-emption

*Intuitive explanation*: A case of Bunzl late pre-emption is a case in which two events C1 and C2 causally over-determine an effect E, but also cause an event B which pre-empts C1 and C2 from causing E *via* events I1 and I2 respectively, in a case of late pre-emption.

*Diagrammatic reconstruction*: (see Lewis [1986b] p. 208)
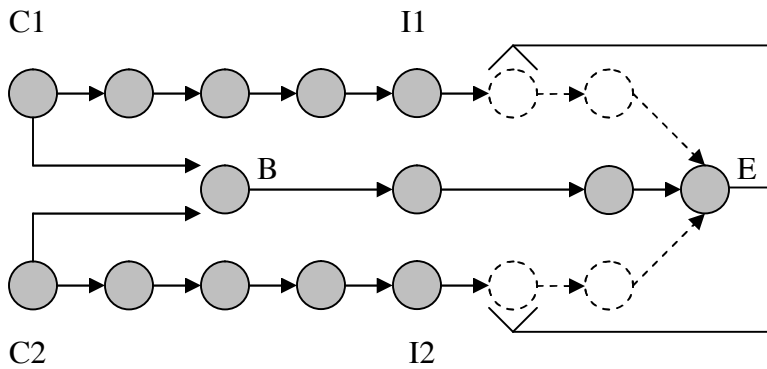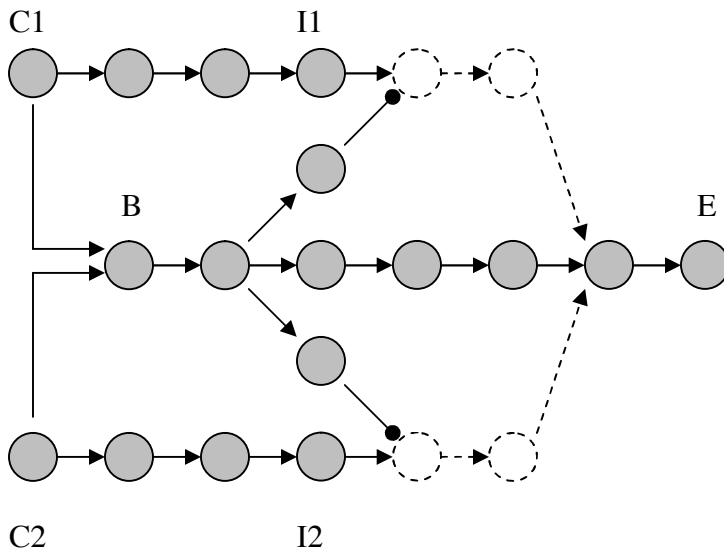


The actual world



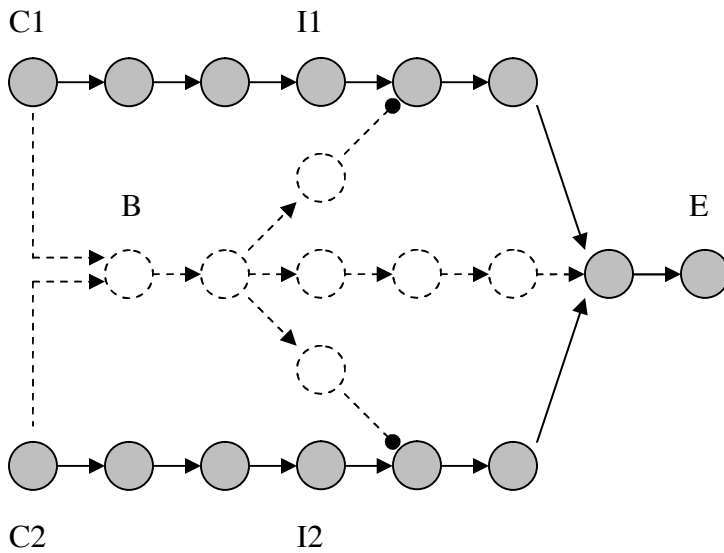The closest ~B world to the actual world

**Example (31)**: NEURONS #1 (an example from Lewis [1986b] p. 208): Suppose two neurons fire (C1 and C2), and that this causes a third neuron to fire (B), where the neuron at event (B) will only fire if stimulated by the firings of both our first two neurons (C1, C2), and that the firing of the neuron at event (B) causes the neuron at (E) to fire, pre-empting neurons firing (I1 and I2) from causing E in a case of late pre-emption. Suppose

it is also the case that if either one of the neurons at (C1) or (C2) had not fired, then the neuron at (B) would not have fired, but the I1 and I2 would then have caused E. So far we have a case of causal over-determination (given both C1 and C2 cause and are causally sufficient for E), and a case of late pre-emption (given the process running from B to E cuts off events I1 and I2 from causing E).

## 3.6.2 Bunzl early pre-emption



The actual world

The closest possible world in which B doesn't occur.

**Example (32)**: NEURONS #2 (an example from Lewis [1986b] p. 208): Suppose two neurons fire (C1 and C2), and that this causes a third neuron to fire (B), where the neuron at event (B) will only fire if stimulated by the firings of both our first two neurons (C1, C2), and that the firing of the neuron at event (B) causes the neuron at (E) to fire, pre-empting neurons firing (I1 and I2) from causing E in a case of early pre-emption. Suppose it is also the case that if either one of the neurons at (C1) or (C2) had not fired, then the neuron at (B) would not have fired, but the I1 and I2 would then have caused E. So far we have a case of causal over-determination (given both C1 and C2 cause and are causally sufficient for E), and a case of early pre-emption (given the process running from B to E cuts off events I1 and I2 from causing E).

173

*Dealing with the examples*: The examples of Bunzl early and late pre-emption are noticeable insofar as both cases include a case of symmetrical over-determination (C1 and C2 are both causes and are both causally sufficient for E), but also contain a case of early/late pre-emption. What is interesting is that the examples are analysable as if they were cases of early and late pre-emption by our analyses that appeal to 'getting rid' or 'holding fixed' strategies, rendering such strategies equally successful here. It should now be obvious how such strategies succeed- we merely 'hold fixed' or 'get rid' of events I1 and I2 in our Bunzl cases for the effect E to be dependent on C2 and C1 independently in the usual way that the holding fixed and quasi dependence analyses respectively permit.

We note that Bunzl's observation does not completely save the counterfactual approach; it is not the case that in *all* cases of symmetrical over-determination a Bunzl event can be identified (and thus there are still cases of pre-emption in which none of the counterfactual analyses we have visited so far provide a complete analysis); indeed, we visited such cases in the section on double pre-emption cases in which no Bunzl event can be guaranteed to be found (in particular, see LIGHT BULB, example 30, p. 161). However, it does show how *some* cases of symmetrical over-determination can be dealt with by counterfactual analyses of causation that were otherwise unfruitful in dealing with the cases of symmetrical over-determination we met earlier.

 I shall now move onto to two final sorts of cases which involve cases of prevention.
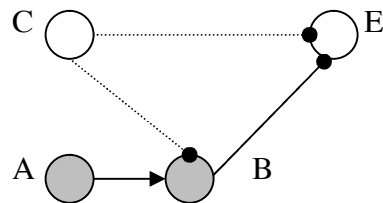
## 3.7 Pre-emptive prevention

*Intuitive explanation*: In cases of pre-emptive prevention, an event C causes the prevention of E and pre-empts B from causing the prevention of E by preventing B itself (fig. 18). Furthermore A would have caused B to prevent E had C not prevented B (fig. 19).[107]

*Diagrammatic reconstruction:*
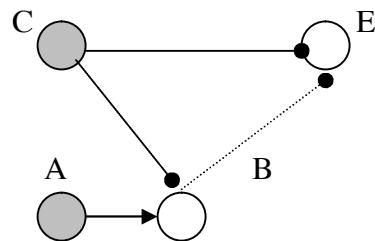


The actual world; C prevents E and B.



---

[107] Note that in examples of this sort omissions feature as causes. Omissions are not events as we have defined them, but might better be described as the absence of certain properties local to space and time. Call then omissions 'negative events' and events as we had originally construed them 'positive events', and thereby allow an analysis with events as the causal relata to include omissions as its causal relata.

The closest possible worlds in which C does not occur; B prevents E

### 3.7.1 Basic pre-emptive prevention

**Example (33)**: SUZY AND SALLY (an example due to Collins[108]): Suppose that Suzy stands in front of a window, and that Sally is standing a short distance away from Suzy and the window. A ball is thrown, and had its flight been uninterrupted (without Suzy or Sally waiting to intervene), it would have shattered the window. Suppose Suzy catches the ball (C), but if Suzy had not, Sally would have jumped in and caught the ball (B). All things being equal, we say Suzy's catch (C) (and not Sally's) prevents the shattering of the window, or, in other words, that Suzy's catch (C) caused the window not to shatter (E).

We shall now visit a contrasting case, before making some general comments.

### 3.7.2 Strengthened pre-emptive prevention

**Example (34):** SUZY AND WALL (a variation of an example due to McDermott[109]): Suppose that Suzy stands in front of a massive brick wall, and the massive brick wall is just in front of a window. A ball is thrown by Billy, and had its flight been uninterrupted (without Suzy or the wall waiting to intervene), it would have broken the window. Suzy catches the ball (C), but if Suzy had not, the presence of the wall (B) would have stopped

---

[108] Collins, [2000], p. 107

[109] See Collins, [2000], p. 107, for some discussion of this case.

the ball from reaching the window; intuitively we think that Suzy's catch (C) does not prevent the shattering of the window (E), nor does the presence of the wall.

*Discussion about the examples*: We should justify our intuitions here. Why is it so clear that Suzy prevented the shattering in the first case and not in the second?

McDermott is one who thinks that Suzy prevents the shattering in both cases, and has the following to say about such cases in defence of this:

> "Suppose I reach out and catch a passing cricket ball. The next thing along in the ball's direction of motion was a solid brick wall. Beyond that was a window. Did my action prevent the ball hitting the window? (Did it cause the ball to *not* hit the window?) Nearly everyone's initial intuition is "No, because it wouldn't have hit the window irrespective of whether you had acted or not." To this I say, "If the wall had not been there and I had not acted the ball would have hit the window. So between us - me and the wall - we prevented the ball from hitting the window. *Which one* of us prevented the ball from hitting the window - me or the wall (or both together)?" And nearly everyone then retracts his initial intuition and says, "Well it must have been your action that did it - the wall clearly contributed nothing."[110]

---

[110] McDermott, [1995a], p. 525.

McDermott would then have the following to say about SUZY AND WALL: We might initially think that Suzy's catch did not cause the prevention of the shattering, because the shattering would have been prevented 'irrespective' of whether Suzy caught the ball or not, given the presence of the wall. But we must eventually yield to the idea that Suzy's catch indeed caused the prevention of shattering, even given the presence of the wall, because the wall's presence 'contributed nothing' to the prevention of the shattering in the actual case. We might suppose McDermott would tell a similar story for SUZY AND SALLY.

Collins is one who disagrees (I think rightly) with this conclusion. He argues that Suzy prevents the shattering in the first case, but not in the second, and thinks there is a subtler story to tell. He argues that in cases where the *wall's* blocking the ball is pre-empted (in example (34)), it seems wrong to say that my catch prevented the shattering. Rather, we find ourselves wanting to say the following things:

- My catch prevented the ball from hitting the wall.

- The wall would have prevented the ball from breaking the window.

- My catch prevented the wall from preventing the ball from breaking the window.

- The wall prevented my catch from preventing the ball from breaking the window.

Collins argues that the relevant difference between the two examples seems to be the following:[111] Given the presence of the wall, the window was never in any danger of

---

[111] Collins, [2000], p. 108-110

being broken.  But in contrast, given the presence of Sally in the first example, the window *was* in such a danger; after all, Sally isn't always a perfect catch! So it seems the wall makes my catch irrelevant to the prevention of the shattering, but that Sally's pre-empted catch does *not* make Suzy's catch irrelevant to the other shattering (it does, on Collins' account, seem to contribute *something* to the prevention of the shattering). And so we are inclined towards the conclusion that Suzy prevented the shattering in the original example, but that her catch did not prevent the shattering in McDermott's.

Collins supports this conclusion by appealing to other cases in which McDermott's treatment of such cases is reduced to absurdity. Consider:

**Example (35):**

COSMIC BASEBALLER (an example due to Collins[112]): Consider a fielder in a baseball field (with no brick wall or window involved), and suppose the fielder makes a catch and that the velocity of the ball points to a vector directed to a point several million miles away from earth, and that by the time the ball could get to this vector, it would collide with Halley's comet on its next-but-one swing through the solar system. Clearly, it seems, my catch did not prevent the ball from colliding with Halley's comet. It is just too 'far-fetched' to suppose that the other factors (such as planet Earth being in the way, between the throw and the comet!) would ever have not made it impossible for the circumstances that enabled the collision to occur.

---

[112] Collins, [2000], p. 110

179

It would be obviously absurd to argue that the baseballer's catch prevented the ball from hitting the comet, but that is precisely the conclusion we must accept if we accept McDermott's reasoning - upon which we would reason "clearly, the intervening events between the catch and the meteor *contributed nothing* to the prevention of the collision of the ball with the comet, and so we must accept the conclusion that the baseballer's catch prevented the comet collision."

Thus, pro Collins and contra McDermott, it seems clear that there are cases that look like cases pre-emptive prevention where we intuitively should allow the pre-empting preventer to be the cause of the prevented event (such as cases where Sally's leap was the only back-up preventer), and cases where we intuitively shouldn't (such as cases where brick walls, or the presence of planets, are the back-up preventers). In the latter sort of cases it seems far more 'far-fetched' to suppose that the catch does any preventing in contrast to the former sort of cases.

So I agree with Collins' conclusions, but have something more to say. I think we should be led to the conclusion that Sally contributed nothing, but the wall in fact did contribute something. The reasons for this are as follows:  Although it is true that the two examples exhibit the same basic structure of counterfactual dependency, and that in both cases the window was not in danger of being smashed, it is also true that in the first example, Sally contributed *nothing* with respect to preventing the shattering *because* the events in which Sally would do some preventing *are not* already in place; her hand would have to *move*

from its original location, some distance away from the trajectory of Billy's ball. In contrast, the wall contributed *something* with respect to preventing the shattering, because the events in which the wall would have prevented the shattering are *already in place*; the wall need do nothing to ensure the prevention of the shattering. Thus, I think the asymmetry of the two cases is located in the fact that the pre-empted preventer in the Sally case is not already an effective and in-place 'blocker' of the shattering, where the pre-empted preventer in the wall case already is an effective and in-place 'blocker'. To generalise, I then say that if C prevents E, then there aren't any in-place 'blockers' already lying in situ that would have prevented E. We shall revisit this sort of reasoning later when I introduce my analysis of causation (chapter 4).

We thus achieve our conclusion that Suzy prevented the shattering in the first case but in not the second. How do our analyses fare up against such examples? Example (33) first:

*The naïve counterfactual analysis and the fine-grained analysis fail*: The naïve counterfactual analysis fails; if Suzy had not caught the ball, the non-shattering still would have occurred, because Sally would have caught the ball in place of Suzy. The fine-grained analysis fails also: the fine-grained analysis saw its early success in being able to appeal to a fine-grained event of the effect. However, in cases of pre-emptive prevention, there is no fine-grained event of the effect, because there is no event of the effect simpliciter, only an omission of the shattering, and so cases of pre-emptive prevention are beyond the scope of the fine-grained analysis. Thus, the naïve

counterfactual analysis and the fine-grained analysis both fail to analyse Suzy's catch (C) as a cause of the non-shattering (E); the incorrect analysis.

*The counterfactual chains analysis fails*: Recall the details of the counterfactual chains analysis;

- C causes E iff there is a chain of counterfactual dependence from E to C.

On the supposition that we may take omissions as well as events as the causal and counterfactual relata, we observe that there is no chain of counterfactual dependence from the non-shattering (E) to Suzy's catch (C). All the chains of counterfactual dependence from Suzy's catch (C) leading to its causal future do not lead to the non-shattering, for example; if Suzy had not caught the ball, then Sally would have, and if Sally would have caught the ball, there still would have been a non-shattering (E). Thus, the counterfactual chains analysis fails to analyse Suzy's catch (C) as a cause of the non-shattering (E); the incorrect analysis.

*The dependence fixing analysis succeeds in example* (33): Recall the details of the dependence fixing analysis:

- C causes E iff there is some event F such that if F had still occurred and C not, then E would not have occurred.

On the proviso that the causal relata may admit omissions, the analysis succeeds in analysing our example, as follows:

1. There is some event F (namely the event which constitutes Sally waiting inertly in the background), such that if F had still occurred and Suzy's catch (C) not, then the non-shattering (E) would not have occurred.

Once again, the strategy of 'holding fixed' troublesome background events comes up trumps. However, the strategy here fails us in example (34), for very strange reasons. Recall in example (34) there is a massive wall in between Suzy and the window which would prevent any ball Suzy fails to catch from reaching the window, in ordinary counterfactual scenarios. The dependence fixing analysis renders Suzy's catch a cause of the non-shattering in a very peculiar manner. As follows:

1. There is some event F (namely the event which constitutes the precise way in which air flowed around the wall, but not above the wall), such that had F still occurred and (C) not occurred, the non-shattering (E) would not have occurred.

What happens when we 'hold fixed' the way in which the air flows just around the wall? We can suppose that holding these events fixed in the right way, but letting go of Suzy's catch, makes the only relatively plausible trajectory of the ball one of over the wall and into the window; smashing it! This is a highly contrived way to obtain counterfactual dependence of the non-shattering (E) on Suzy's catch (C). The basic idea to the problem

is that the dependence fixing analysis allows us to hold fixed (sometimes very obscure) events that force a very unnatural causal path for the remaining active causal processes.

*All the variations of the influence analysis fail*: We shall only demonstrate that the influence analysis fails for example (33) here, given it should now be obvious how its variations will fail. Recall the details of the influence analysis:

- C causes E iff C influences E.
- C influences E iff there is a substantial range of $C_1 \ldots C_n$ of different not-too-distant alterations of C (including the actual alteration of C) and a range of $E_1 \ldots E_n$ of alterations of E, at least some of which differ, such that if $C_1$ had occurred, $E_1$ would have occurred, if $C_n$ had occurred, $E_n$ would have occurred.[113]

Here we must revise the influence analysis in order to admit omissions (such as non-shatterings) into the causal relata, but doing so does not help us with regard to the examples of pre-emptive prevention, for there is no sense in which non-too-distant alterations to Suzy's catch (i.e. alterations on how she catches, when she catches, and whether she catches) counterfactually imply alterations to the non-shattering (i.e. on how the window doesn't shatter, when it doesn't shatter, and whether it doesn't shatter). For starters, we must ask what an alteration to a window's 'non-shattering' would even constitute, and even whether such a locution could make sense. In any case, and at least

---

[113] Lewis, [2003], p. 91

as far as sense can be made of such locutions, we may conclude that non-too-distant alterations to the timing or manner of Suzy's catch (C) do not alter the non-shattering (given the same 'non-shattering' would have still occurred), nor do alterations to whether Suzy's catch (C) occurred or not (given Sally would have jumped in to catch the ball in place of Suzy). Thus we derive the conclusion that Suzy's catch (C) does not cause the non-shattering (E); the incorrect analysis.

*The would-be dependence analysis succeeds in example (33)*: Recall the details of the would-be dependence analysis:

- C causes E iff there is a chain of counterfactual dependence linking C to E, or there would be such a chain had it not been for some pure dependence preventer, and the would-be chain still occurred

- An event P purely prevents dependence of E on C iff P is a dependence preventer on a chain of events from E to C, and P prevents no event in a chain of events that would form a counterfactual chain linking E to C had P not occurred.

- P is a dependence preventer of E on C iff had P not occurred and the C-E chain of events still had, then E would counterfactually depend on C.

The would-be dependence analysis succeeds in analysing Suzy's catch (C) as a cause of the non-shattering (E), as follows:

- The non-shattering (E) would counterfactually depend on Suzy's catch (C), had it not been for some pure dependence preventer (namely, Sally's presence) and the non-shattering (E) and Suzy's catch (C) still occurred.

*The would-be dependence analysis fails in examples (34) and (35)*: The problem with the would-be dependence analysis, as it has just been stated, is that the strategy of 'getting rid' of background events over-generates causes in examples (34) and (35). Take example (34), in which there is a massive wall in position in between Suzy and the window. In this example, we can use the would-be dependence analysis to analyse Suzy's catch (incorrectly) as a cause of the non-shattering, given:

- The non-shattering (E) would counterfactually depend on Suzy's catch (C), had it not been for some pure dependence preventer (namely, the wall's presence) and the non-shattering (E) and Suzy's catch (C) still occurred.

Collins has a response to this problem. He suggests that we revise the would-be dependence analysis as follows:

- C causes E iff there is a chain of counterfactual dependence linking C to E, or there would be such a chain had it not been for some pure dependence preventer *in some not-too-far-fetched way*,

The essential idea is that we can only 'get rid' of troublesome pre-empted back-up causes if getting rid of them wouldn't be too far-fetched. Collins then argues that getting rid of the wall in example (34) would be 'too far-fetched', and Suzy's catch would thereby resist being analysed as a cause.

Collins' proviso is of course inadequate. Firstly, it is left unclear as to what exactly constitutes 'far-fetched' in a way that doesn't render the analysis of causation viciously circular; and secondly (and more importantly), there seems no way to cash out the notion in a non circular fashion that would seem to adequately do the job. For example, 'not too far-fetched' can't mean 'not too distant', given there may be pre-empted back-up causes which we have to get rid of, but can only do so by appeal to very distant worlds. Collins' revised analysis should only be accepted if he can cash out the notion of what he takes the notion of 'far-fetched' exactly to imply, and in a way to do the job he intends.

*The quasi-dependence analysis fails in example* (33)*:* Recall the details of the quasi-dependence analysis:

- C causes E iff there is an isonomic duplicate of a process from C to E, in which the duplicate of E counterfactually depends on the duplicate of C.

The analysis fails to analyse Suzy's catch (C) as a cause of the non-shattering (E), simply because there is no process (of the intuitive sort) from C to E at all! Example (33) thus serves to expose the quasi-dependence analysis as incomplete; there are cases of

187

causation that do not involve processes in the way the quasi-dependence analysis seems to require (see Lewis [2001a]) in order to get off the ground. The analysis' incompleteness can be remedied by narrowing the scope of the analysandum, namely by making the right hand side of the bi-conditional an analysis of "C causes E via a process", instead of causation in general. However, this still leaves the question of finding an analysis of causation in general an open one. I conclude this section by tabulating some results:

*Some observations*: We catalogue the projected and demonstrated results for each analysis as they deal with the examples:

- The Humean counterfactual analysis, the fine-grained analysis, and counterfactual chains analysis fail in all cases
- The holding fixed analysis succeeds in all cases
- The influence analyses fail in all cases.
- The would-be dependence analysis succeeds in 33, but fails in 34 and 35
- The quasi-dependence analysis fails in all cases.

Once again, we note that the strategy of 'holding fixed' troublesome pre-empted events, has the most success. However, the strategies of 'getting rid' of background events (such as Collins' would-be dependence analysis, and Lewis' quasi-dependence analysis), which had a lot of success in other cases of pre-emption, have made notable failures. The main problem of the 'getting rid of' strategy (as appealed to by the would-be dependence

analysis) is that it over generates causes; getting rid of some background events can promote another set of events to the status of being causal, when in fact no such causation is occurring. The main problem of the 'getting rid of' strategy (as appealed to by the quasi dependence analysis) is that it under generates causes; sometimes there is no process that the analysis can appeal to from the cause to the effect in order for the analysis to succeed. These sorts of problems are in no way minor- to see the full damage that 'getting rid of' strategies incur with respect to cases of prevention; we visit cases of double prevention.

## 3.8 Double prevention

*Intuitive explanation*: In cases of double prevention, A and C cause E. C prevents D, and the prevention of D prevents E from being prevented. If C had not occurred, then B would have caused D and D would have prevented E.
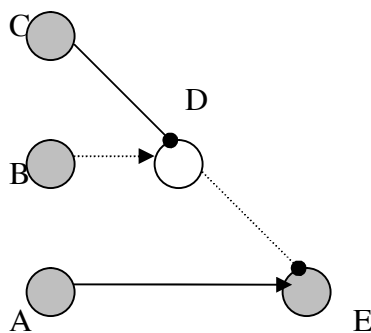


The actual world (W0): C and A occur.
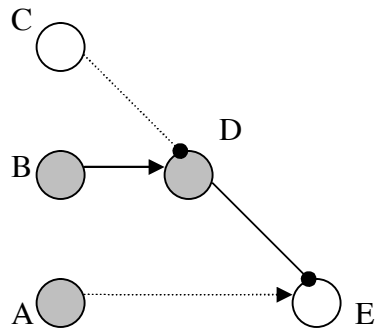
The closest possible world to W0 in which C does not occur.

**Example (36)** BOMBER (an example due to Hall[114]): Suppose Suzy is piloting a bomber (A) on a mission to blow up an enemy factory, and Billy is escorting her in his fighter plane (C). An enemy fighter plane attempts to destroy Suzy's bomber (B) in order to prevent her from completing her mission, but Billy manages to destroy the enemy plane just before the enemy manages to destroy Suzy's bomber (D). Suzy's mission thus goes uninterrupted, and the bombing occurs as planned (E). However, if Billy had not been escorting Suzy (C), he would not have shot down the enemy (D), and the enemy would have shot down Suzy's bomber, and in turn the blowing up of the factory (E) would have been prevented. Thus, Billy's escorting Suzy (C), prevents the enemy from shooting down Suzy's bomber (D), which in turn prevents the enemy plane from preventing Suzy's bombing (E).

---

[114] p. 241, 'Two Concepts of Causation' in Collins, Hall, and Paul.

**Example (37)** PRESIDENT (an example due to McDermott [1995a]): Suppose a crazed President is about to launch a nuclear strike (B), and that luckily, the President's sensible assistant prevents him from doing so (C). Suppose Joe Blow is feeling hungry (A), which causes him to have breakfast (E). Suppose also that if the President had launched a nuclear strike, then Joe Blow would have been blown up just before he was able to have breakfast (E). Thus, Joe Blow's hunger (A), was a cause of his eating breakfast, but so is the assistant's preventing the president from launching a nuclear strike (C).

**Example (38)** BILLIARDS (an example due to Lewis[115]): Suppose the collision between billiard balls 1 & 2 (C) prevents the collision between balls 1 & 3 (D), which prevents the collision of balls 3 & 4 from being prevented (E). So suppose that if the first collision (C) had not occurred, ball 1 would have taken a new trajectory that would have caused (D), and that the trajectories of ball 3 & 4 (A) are a cause the second collision (E) in the actual world.

*Topographical diagram of example 38:* For an illustration of this example see the topographical diagram below: Grey circles represent billiard balls, black arrows represent their trajectories, white circles represent collisions as the result of the billiard balls' trajectories, a dashed circle represents a collision between 1 & 3 that would have occurred had the collision between 1 & 2 not occurred, dashed arrows represent their trajectories if the first collision had not occurred. We ignore the strongly dashed vertical line marked 'W' in the middle for now).

---

[115] Lewis, 'Causation as Influence', [2000], p. 83-4.
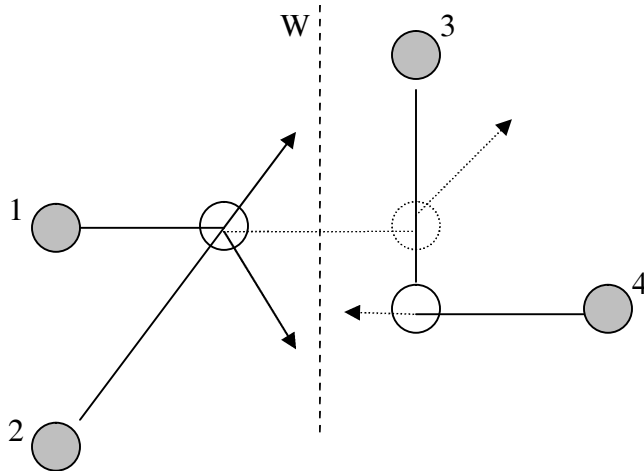
*Fig\*,* A topographical diagram of Lewis' example.

*Dealing with the cases***:**

According to Lewis, in cases of double prevention both (C) and (A) should be analysed as causes of E.[116] In the first case, Suzy's bomber (C) causes the destruction of the factory, but so does Billy's fighter destroying the enemy (A). In the second case, Joe Blow's hunger (A) is a cause of his eating breakfast, but so is the advisor stopping the President from starting a nuclear war (C). In the third case, the trajectories of billiard balls 3 and 4 cause the collision, but so does the collision of balls 1 and 2. How do our analyses fare up against such like cases?

---

[116] Lewis, 'Causation as Influence', [2000], and 'Void and Object', [2000].

192

The answer is that all of our prominent analyses are successful, apart from those analyses that appeal to a 'getting rid of' strategy (i.e. the quasi-dependence analysis and Collins' would-be dependence analysis). The failure of the quasi-dependence analysis comes easily - there is no process (of the intuitive sort) from event A to E in our examples, and thus no causation overall; the incorrect analysis. The failure of the would-be dependence analysis comes from making a revision to the sorts of double prevention cases we have been considering. Consider fig.* again, except this time suppose that the long dotted vertical line marked W is a brick wall that would effectively stop any ball on one side of the wall from reaching the other side. In this case, the collision of balls 1 and 2 does not cause the collisions of balls 3 and 4 (this is because it is no longer the case that if the collision between 1 and 2 hadn't occurred the collision of 3 and 4 wouldn't have, given the wall in between would have prevented balls 1 and 2 from interfering with the collisions of balls 3 and 4).

**3.9 Lessons and Morals for a successful counterfactual analysis of causation**:

In this section I wrap up some general observations and morals from our investigations.

So far I have exposed some of the most prominent analyses of causation to date, and then tested them against the deterministic pre-emption examples, and despite some successes (which have been catalogued), there has been not one analysis that can handle all of the cases. We should then here make some general observations about the strategies we have

investigated so far, as doing so will give us helpful clues as to what a successful counterfactual analysis of causation will look like.

1. *The appealing to 'chains' strategy.* A strategy that Lewis employed for all three of his analyses was to appeal to chains of dependence, quasi-dependence, or influence in order to deal with the pre-emption cases. As we have seen, not only was this generally unsuccessful (and unsuccessful on a massive scale) with respect to handling such cases, but the strategy bore the cost of restricting what sort of causal concept we were analysing from the very start; that is, the analysis only allowed us to analyse a transitive causal concept of 'causal histories' as opposed to a non transitive causal concept more familiar to that of 'causal explanation'.

2. *The appealing to 'fragile' effects strategy.* Another strategy was to appeal to a fragile version of the effect in our analyses themselves. As we have seen, not only was this strategy generally unsuccessful (and unsuccessful on a massive scale) with respect to handling pre-emption cases, but the strategy bore the cost of restricting what sort of causal concept we were analysing from the very start; that is, the analysis only allowed us to analyse causation between fine-grained entities, as opposed to causation between coarse-grained entities.

3. *The appealing to 'influence' strategy.* Another strategy was to appeal to Lewis' notion of influence, which took its strength from noticing that effects co-vary with their causes. As we have seen, not only was this strategy generally unsuccessful (and unsuccessful on

a massive scale) with respect to handling the pre-emption cases, but the strategy bore the cost of restricting what sort of causal concept we were analysing from the very start; that is, the analysis only allowed us to analyse a concept of 'causal influence' as opposed to causation simpliciter, and additionally proved unfruitful when dealing with causation by and of omissions.

4. *The appealing to 'processes' strategy.* Another strategy was to appeal to an unanalysed notion of processes that linked cause and effect in order to deal with the pre-emption cases (namely, in Lewis' quasi-dependence analysis). As we have seen, the strategy bore the cost of being unable to analyse causation that held between events or omissions that were not clearly linked by a process; a problem provided by the fact that we were furnished with no clear truth conditions for what constituted a process. That said, if an analysis of processes became available (that was additionally adaptable for causation that included omissions), the strategy could be a profitable one.

5. *The 'getting rid of' strategy.* Another strategy was to appeal to a strategy of 'getting rid' of the troublesome pre-empted causes that prevented counterfactual dependence, and then testing for dependence. This was one of the most successful strategies, although it was unable to deal with *all* cases of pre-emption when the quasi- and would-be dependence analyses employed it.

6. *The 'holding fixed' strategy.* Another strategy was to appeal to a strategy of 'holding fixed' the troublesome pre-empted causes that prevented counterfactual dependence, and

then testing for dependence. This was one of the most successful strategies, although it was unable to deal with *all* cases of pre-emption when the dependence and influence fixing analyses employed it.

7. *The 'embedded counterfactual'* strategy. Another strategy was to appeal to a strategy of employing embedded counterfactuals. This strategy was employed by the would-be dependence analysis, where the idea was not to ask whether the effect counterfactually depends on the cause, but to ask whether the effect *would* counterfactually depend on the cause had circumstances been altered in a systematic way. This was also one of our most successful strategies.

These observations suggest some very important morals for a successful counterfactual analysis of causation, as follows:

- The analysis should be easily revisable to accommodate a non-transitive and transitive relation (from 1)
- The analysis should be easily revisable to accommodate both causation between fine and coarse grained events, and should allow omissions as the relata (from 2)
- The analysis, if it appeals to processes in its analysis, should provide clear truth conditions for what indeed constitutes a process (from 4)
- The analysis might do well to take advantage of 'getting rid of' and 'holding fixed' strategies (from 5 & 6)

- The analysis might do well to employ a strategy of 'embedded counterfactuals' (from 7)

- The analysis has a missing element, the intuition being that this missing element has something to do with counterfactuals.

In the next section, I shall take each of these morals by the horns, and develop a counterfactual analysis of causation that can successfully deal with all the pre-emption cases we have visited. After introducing my analysis, I am going to (very briefly!) re-present the core types of pre-emption examples we visited earlier, and then show how my analysis deals with them. I have found this forthcoming brief re-presentation unavoidable for two reasons: firstly, I am going to have go into more detail with respect to the examples in order to clearly show how my analysis deals with them (something which would have only superfluously convoluted matters when I investigated such cases earlier on); and secondly the re-presentation renders it a lot easier to understand how my analysis deals with the cases (for the weary reader this shall not be too taxing; the sum of the examples I give only amounts to 1000 words).

**Chapter 4: My Analysis; The 'Ps and Qs' analysis**


**4.1 A Counterfactual Analysis of Deterministic Causation, and Pre-emption Cases.**

# 1 The analysis

## 1.1 Causation

Over the past few decades counterfactual analyses of causation have proliferated in almost immeasurable abundance,[117] and with two things in common; firstly, they make much of counterfactual dependence, and secondly, none of them successfully handles all the pre-emption cases. In response to the continuing debate, I suggest that counterfactual *dependence* is only one side of the coin with respect to a successful analysis of causation, and that the notion of counterfactual *implication* has an equal role to play in dealing with the pre-emption cases. Drawing from influences in the literature, I analyse causation as a matter of processes and background factors that exhibit the right pattern of would-be counterfactual dependence and implication, and present a complementary analysis of processes (part 1). I then show how my analysis can successfully handle all the six deterministic pre-emption cases (part 2). I then conclude with some morals for the ongoing debate (part 3).

What is causation? Paradigm cases of causation elicit a pattern of would-be counterfactual dependence; if the cause and effect had occurred, then it would have been the case that if the cause hadn't occurred then the effect wouldn't have. This sort of fact

---

[117] For some prominent accounts, see Collins [2000], Lewis [1973b], [1986c], [2001a], Noordhof [1999], Ramachandran [1997], and Yablo [2002].

has been well noted in the literature.[118] However, what has been ignored up until now is that paradigm cases of causation *also* elicit a pattern of would-be counterfactual implication; if neither the cause nor effect had occurred, then it would have been the case that if the cause had occurred then the effect would have[119],[120]. An analysis of causation in paradigm cases then suggests itself; where C and E are distinct actual events, then:

Paradigmatically, C causes E iff

      1. $(o\text{C} \,\&\, o\text{E}) \,\square\!\!\rightarrow\, (\sim\!o\text{C} \,\square\!\!\rightarrow\, \sim\!o\text{E})$

      2. $(\sim\!o\text{C} \,\&\, \sim\!o\text{E}) \,\square\!\!\rightarrow\, (o\text{C} \,\square\!\!\rightarrow\, o\text{E})$[121]

---

[118] Starting with Lewis [1973b] to the present day (see references).

[119] Following Lewis [1973a], a counterfactual 'if C had occurred then E would have occurred', symbolised '$(o\text{C} \,\square\!\!\rightarrow\, o\text{E})$', is true iff the closest (i.e. most similar) worlds in which C occurs are worlds in which E occurs. Moreover, world 1 (w1) is more similar to world 0 (w0) than world 2 (w2) iff as a matter of 1st priority, w1 avoids *widespread* violations of w0's laws more than w2. As a matter of 2nd priority, w1 *maximises* the spatio-temporal perfect match of particular events from w0 more than w2. As a matter of 3rd priority, w1 avoids *local* violations of w0's laws more than w2. As a matter of 4th priority, w1 secures *approximate* similarity of w0's particular events more than w2.

[120] I adopt a strong notion of negation, such that expressions of the form '$\sim\!o\text{A}$' are defined in terms of classical negation as 'neither A nor any part of A occurs'.

[121] 1. If C and E had occurred, then it would have been the case that if C hadn't occurred then E wouldn't have occurred, and 2. If neither C nor E had occurred, then it would have been the case that if C had occurred then E would have occurred.

The analysis is intuitive; where would-be counterfactual dependence reflects the round-about 'necessity' of causes for their effects (condition 1), would-be counterfactual implication reflects the round-about 'sufficiency' of causes for their effects (condition 2). Take a paradigm case: My throw (C) causes the shattering (E) of the window. Why so? Two counterfactuals are true: 1. If my throw (C) and the shattering (E) had occurred (which they do), then it would have been the case that if my throw (C) hadn't occurred then the shattering (E) wouldn't have. 2. If neither my throw nor the shattering had occurred, then it would have been the case that if my throw (C) had occurred then the shattering (E) would have. Thus, my throw (C) is a cause of the shattering (E).

Unfortunately, however, the analysis fails to analyse non-paradigmatic examples of causation, namely, cases of redundant causation. Typical redundant causes may be characterised as follows (see Lewis [1986c], p. 193). Where C, Q, and E are distinct actual events,[122] then:

Typically[123], Q is a redundant cause of E iff

    1. $(o\text{Q} \ \& \sim o\text{C}) \ \square\!\!\rightarrow o\text{E}$

    2. $(\sim o\text{Q} \ \& \ o\text{C}) \ \square\!\!\rightarrow o\text{E}$

---

[122] Following Lewis [1986b], we may assume that events are intrinsic properties local to space and time. I leave it open as to whether events are fine-grained (in which case all their properties are essential to them) or coarse-grained (in which case some (but not all) of their properties are essential to them).

[123] Contra Lewis, this isn't a complete characterisation of redundant causes; cases of middle pre-emption evade the Lewisian analysis.

3. $(\sim o\text{Q} \& \sim o\text{C}) \; \square\!\!\rightarrow \; \sim o\text{E}$[124]

Here's a typical case of redundant causation: Suppose my throw (C) of a rock causes the shattering (E) of a window, but that if my throw (C) hadn't occurred, your throw (Q) would have caused the shattering (E), although in fact your rock never reaches the window. In this case, the paradigmatic analysis fails to yield the correct result that my throw (C) causes the shattering (E): although the second condition of the paradigmatic analysis is satisfied (if neither my throw (C) nor the shattering (E) had occurred, then my throw (C) would counterfactually imply the shattering (E), given my throw (C) was sufficient for such a shattering (E)), the first condition isn't (if my throw (C) and the shattering (E) had occurred, then the shattering (E) *wouldn't* counterfactually depend on my throw (C), given your throw (Q) would have sufficed for the shattering (E) in place of mine). Thus my throw (C) is incorrectly analysed as not a cause of the shattering (E). What to do? Dropping only one of the formal conditions in our paradigmatic analysis does us no good; where the first alone under-generates causes (it fails to analyse pre-empting causes as causes), the second alone over-generates causes (it analyses pre-empted non-causes as causes).

The paradigmatic analysis thus requires amendment. How to approach this? To return to the last example, we might begin by noticing that what distinguishes my throw (C) from yours is the existence of a process from my throw to the shattering (E), where there is no

---

[124] 1. If Q had occurred and C hadn't, then E would have occurred. 2. If Q hadn't occurred and C had, then E would have occurred. 3. If neither C nor Q had occurred, then E wouldn't have occurred.

process leading from your throw (Q) to the shattering (E). Two counterfactuals are then true: 1. If the process of my throw (C) to the shattering (E) had occurred without the background factor of your throw (Q), then the shattering (E) would counterfactually depend on my throw (C) (given your throw would no longer be present to prevent dependence of the shattering (E) on my throw (C)). 2. If the process of my throw (C) to the shattering (E) hadn't occurred but the background factor of your throw (Q) had, then my throw (C) would counterfactually imply the shattering (E) (even in the presence of your throw (Q)). This suggests an analysis of causation:

C causes E iff there is a process P of C to E, and some background factors Q such that:

    1. $(o\text{P} \,\&\, \sim o\text{Q}) \,\square\!\!\rightarrow (\sim o\text{C} \,\square\!\!\rightarrow \sim o\text{E})$

    2. $(\sim o\text{P} \,\&\, o\text{Q}) \,\square\!\!\rightarrow (o\text{C} \,\square\!\!\rightarrow o\text{E})$[125]

We let our background factors Q admit any collection of actual events distinct from P, or just the omission of P[126]. This is my analysis of causation, which I spell out in more detail shortly. The strategy contains elements both old and new: Condition 1 uses a strategy of

---

[125] 1. If the process P had occurred and the background factors Q hadn't, then it would have been the case that if C hadn't occurred E wouldn't have, and 2. If the process P hadn't occurred and the background factors Q had, then it would have been the case that if C occurred then E would have. (Note that the factivity of this analysis will be implied by the analysis of processes I offer shortly.)

[126] Letting Q admit the omission of P is a caveat that permits analysis where there are no background events distinct from a process of C to E at all, and permits analysis in cases where the removal of background events is unnecessary for the effect to counterfactually depend on the cause.

'getting rid' of troublesome background events (like pre-empted back-up causes), 'holding fixed' the process of cause to effect, and then testing for counterfactual dependence of the effect on the cause. Condition 2 uses a complementary strategy of 'getting rid' of the process of cause to effect, 'holding fixed' the troublesome background events (like pre-empted back-up causes), and then testing for counterfactual implication from cause to effect. The former strategy is old; with proponents such as Collins [2000] and Lewis [1986c] differing in the details. The latter strategy is entirely new, and will play a decisive role.

A first question must be whether we can give a non-causation-presuming analysis of processes given their central role in the above analysis. The answer is that we can, and it's to this issue I turn.

## 1.2 Processes

What are processes? Paradigmatically, a process is a chain of actual events (C, D,…, Dn, E), some of which are distinct[127]. It's also a chain of events which exhibits steps of would-be counterfactual dependence; had the chain (C, D,…, Dn, E) occurred, then there would be a corresponding chain of counterfactual dependence from C to E (i.e. if C hadn't occurred, D wouldn't have, if D hadn't occurred, D1 wouldn't have… etc.). It's

---

[127] This provision permits of the indistinctness of cause and effect just when there is a process connecting them, such as in closed causal loops. The causal relation is then correctly rendered non-transitive, non-symmetric, and non-reflexive.

also a chain of events which exhibits steps of would-be counterfactual implication; had nothing on the chain (C, D,…, Dn, E) occurred, then there would be a corresponding chain of (proper[128]) counterfactual implication from C to E (i.e. if C had occurred, D would have, if D had, D1 would have… etc.). But it seems these chains of counterfactuals should afford an extra property; they should be 'complete' in the right way. That is, any event on any chain of counterfactual dependence/implication from C to E in our counterfactual scenarios should also feature (as a counterpart to a part of the sum of the events) on the chain (C, D,…, Dn, E). This seems true enough; we don't want to omit any stage of the process from its start to finish upon its characterisation. An analysis of processes in the paradigm case then suggests itself:

Paradigmatically, P is a process of C to E iff P is a chain of actual events (C, D,…, Dn, E), some of which are distinct, such that:

1. ($o$C, $o$D,…, $o$Dn, $o$E) □□ there is a correspondingly complete chain of counterfactual dependence (~$o$C □□ ~$o$D, ~$o$D □□…,….,…□□ ~$o$Dn, ~$o$Dn □□ ~$o$E).
2. (~$o$C, ~$o$D,…, ~$o$Dn, ~$o$E) □□ there is a correspondingly complete chain of (proper) counterfactual implication ($o$C □□ $o$D, $o$D □□…,….,…□□ $o$Dn, $o$Dn □□ $o$E).

Some details: A chain of counterfactual dependence (~$o$C □□ ~$o$D, ~$o$D □□…,….,…□□ ~$o$Dn, ~$o$Dn □□ ~$o$E) corresponding to a chain (C, D,…, Dn, E) is complete iff any event in any chain of counterfactual dependence from C to E features on the chain (C, D,…,

---

[128] A counterfactual is proper iff it has a false antecedent.

Dn, E). A chain of (proper) counterfactual implication ($o$C □□ $o$D, $o$D □□…,…,…□□ $o$Dn, $o$Dn □□ $o$E) corresponding to a chain of (C, D,…, Dn, E) is complete iff any event in any chain of (proper) counterfactual implication from C to E features on the chain (C, D,…, Dn, E). An event features on the chain (C, D,…, Dn, E) iff that event is a counterpart to a part of the sum of events on the chain.

Take a paradigm case: I throw a rock (C); it travels through early (D1) to mid (D2) to late (D3) flight, and the window shatters (E). Those events make up a process. Why so? There is a chain of actual events P (C, D1, D2, D3, E) such that: 1. If the chain (C, D1, D2, D3, E) had occurred, then it would have been the case that there is a corresponding chain of dependence from my throw (C) to the shattering (E); if my throw (C) hadn't occurred, then the rock's early flight (D1) wouldn't have, if the rock's early flight (D1) hadn't occurred, then its mid-flight (D2) wouldn't have, … etc. This counterfactual chain is 'complete' in the right way; any event in any chain of dependence from C to E is a counterpart to a part of the sum of the chain (C, D1, D2, D3, E). 2. If nothing on the chain (C, D1, D2, D3, E) had occurred, then it would have been the case that there is a corresponding chain of (proper) counterfactual implication from my throw (C) to the shattering (E); if my throw (C) had occurred, then the rock's early flight (D1) would have, if the rock's early flight (D1) had occurred, then the rock's mid-flight (D2) would have, … etc. We note this counterfactual chain is 'complete' in the right way; any event in any chain of counterfactual implication from C to E is a counterpart to a part of the sum of the chain (C, D1, D2, D3, E). Thus, there is a process P consisting in the chain of events (C, D1, D2, D3, E) of my throw (C) to the shattering (E).

Unfortunately, however, this analysis fails to account for non-paradigmatic processes that feature in cases of redundant causation; cases where one step in the process is redundantly caused. To fix this problem, a similar solution suggests itself for processes as it did for causation, namely by a special appeal to background factors in our analysis. We then re-analyse processes as follows:

P is a process of C to E iff P is a chain of actual events (C, D,…, Dn, E), some of which are distinct, such that for some background factors Q:

1. (*o*C, *o*D,…, *o*Dn, *o*E & ~*o*Q) □□ there is a correspondingly complete chain of counterfactual dependence (~*o*C □□ ~*o*D, ~*o*D □□…,….,…□□ ~*o*Dn, ~*o*Dn □□ ~*o*E).

2. (~*o*C, ~*o*D,…, ~*o*Dn, ~*o*E & *o*Q) □□ there is a correspondingly complete chain of (proper) counterfactual implication (*o*C □□ *o*D, *o*D □□…,….,…□□ *o*Dn, *o*Dn □□ *o*E).

Again we let our background factors Q admit any collection of actual events distinct from P, or just the omission of P. This analysis seems to do the job.[129] Together, my analyses are best illustrated by appeal to cases, which we'll visit soon. For now, we note that the analysis of processes forms a nice counterpart to the analysis of causation; both tie in with the spirit of would-be counterfactual dependence and implication. Why not allow an

---

[129] Cases of causation of, by, and via omission are also possible (see, Lewis [2001b], Schaffer [2000], and Beebee [2004] for discussion). For simplicity's sake I ignore this type of causation here and focus on causation between events. That said, the analysis may easily be revised to allow omissions.

analysis of causation in terms of processes alone; i.e. C causes E iff there is a process of C to E? The answer is that such an analysis renders the causal relation invariably transitive. Counter-examples to the transitivity of causation are common-place; Billy's being fat caused him to go on a diet, and his going on a diet caused him to be thin, but his being fat didn't cause him to be thin. Examples like this form part of the folklore of counter-examples to the transitivity of causation (See Hall [2000]), and here my analysis of causation gives the right results; if the process of Billy's being fat to his being thin had still occurred (and for any Q, Q hadn't), then it *wouldn't* be the case that if Billy hadn't been fat he wouldn't have been thin - indeed, if he hadn't been fat he would have been thin.

## 2 The pre-emption cases

Cases of redundant causation prove troublesome for counterfactual-style analyses of causation. In the remainder of this chapter, I set out every type of deterministic pre-emption case discussed in the literature, and show how my analysis can handle each of these. We begin with the cases of early, late, middle, and double pre-emption.

### 2.1 Early pre-emption

Cases of early pre-emption are cases in which the main process running from cause to effect cuts off the pre-empted cause from causing the effect *before* the main process from cause to effect comes to completion (see Lewis [1973b]).
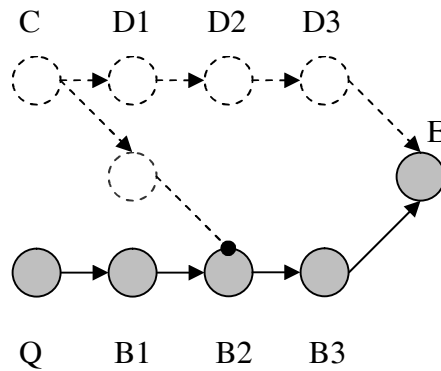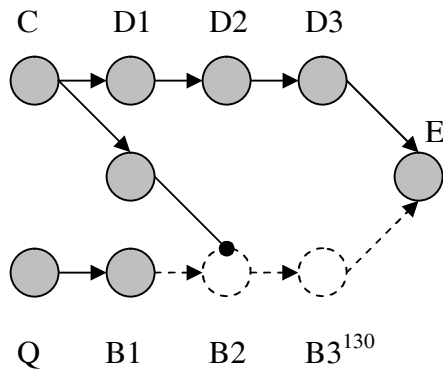
*Example*: Suppose Suzy and Billy are playing with rocks near a window. Suzy throws her rock (C); it flies through the air from early (D1) to mid (D2) to late (D3) flight, and shatters the window (E). Suppose Billy stands next to Suzy and nearly throws (Q) at the window at the time of Suzy's throw (C), but that Suzy's throw (C) prevents Billy from throwing his rock. In short, Suzy's throw (C) (and not Billy's near-throw (Q)) causes the shattering (E), but if Suzy's throw (C) hadn't occurred then the shattering (E) still would have, given Billy would have made a throw sufficient for the shattering (E).

## 2.2 Late pre-emption

---

[130] Grey outlined circles represent actual events, white outlined circles represent actual omissions; their dashed counterparts represent non-actual events/omissions. Black arrows represent actual causal connections, black blobbed lines represent actual causal connections of prevention; their dashed counterparts represent non-actual causal connections. A long black splitting line represents a fact which inhibits an event from causing E; its dashed counterpart represents a fact which would (but actually doesn't) inhibit an event from causing E. A big black arrow represents a 'trumping' causal connection; its dashed counterpart represents a non-actual 'trumping' causal connection.

Cases of late pre-emption are cases in which the main process running from cause to effect cuts off the pre-empted cause from causing the effect *after* the main process from cause to effect comes to completion (see Lewis [2001a]).
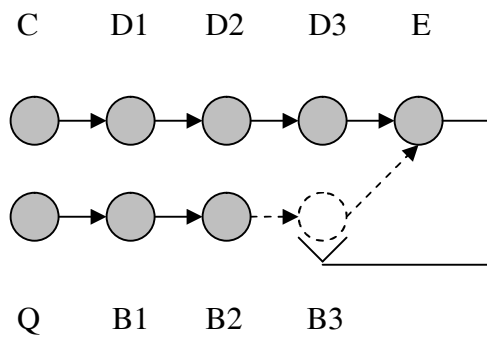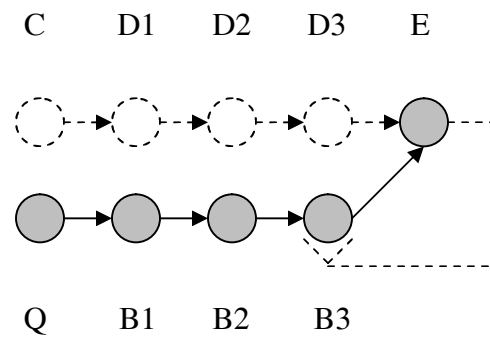
Fig. 3: The actual world (w0)          Fig. 4: The closest ~C worlds to w0



*Example*: Suppose Suzy and Billy are playing with rocks near a window. Suzy throws her rock (C); it flies through the air from early (D1) to mid (D2) to late (D3) flight, and shatters the window (E). Suppose Billy throws his rock (Q) simultaneously with Suzy's throw, and that his rock flies through the air in early (B1) to mid (B2) flight, but fails to forcefully meet the window (B3); given Suzy's rock (which was thrown slightly harder) shatters the window just before Billy's was able to. In short, Suzy's throw (C) (and not Billy's (Q)) causes the shattering (E), but if Suzy's throw (C) hadn't occurred, the shattering (E) still would have, given Billy's throw (Q) would have sufficed for the shattering (E).

## 2.3 Double pre-emption

In cases of double pre-emption, each of two distinct events causally over-determines an effect, or in other words, they cause and are causally sufficient for an effect (see McDermott [1995]).

Fig. 5: The actual world (w0)                     Fig. 6: The closest ~C worlds to w0
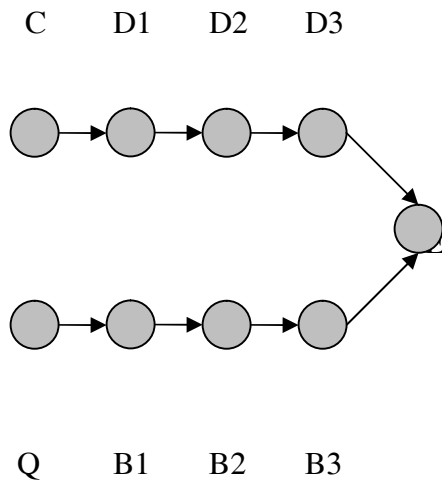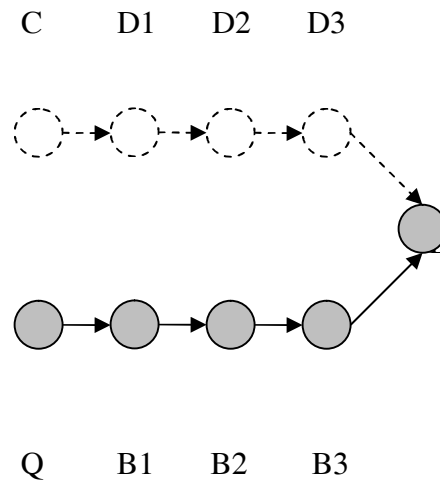


*Example***:** Suppose Suzy and Billy are playing with rocks near a window. Suzy throws her rock (C); it flies through the air from early (D1) to mid (D2) to late (D3) flight, and shatters the window (E). Simultaneously, Billy throws his rock (Q), it flies through the air from early (B1) to mid (B2) to late (B3) flight, and shatters the same window (E) at the same time as Suzy's rock does. In short, Suzy's throw (C) is a cause of the shattering (E), as is Billy's (Q). However, if Suzy's throw (C) hadn't occurred the shattering (E) still

would have, given Billy's throw would have sufficed for the shattering (E), and if Billy's throw (Q) hadn't occurred the shattering (E) still would have, given Suzy's throw (C) would have sufficed for the shattering (E).

## 2.4 Middle pre-emption

In cases of middle pre-emption, some event mid-process from the main cause to an effect results in the pre-emption of another cause (see Strevens [2003]).[131]

Fig 7: The actual world (w0)    Fig 8: The closest ~C worlds to w0



---

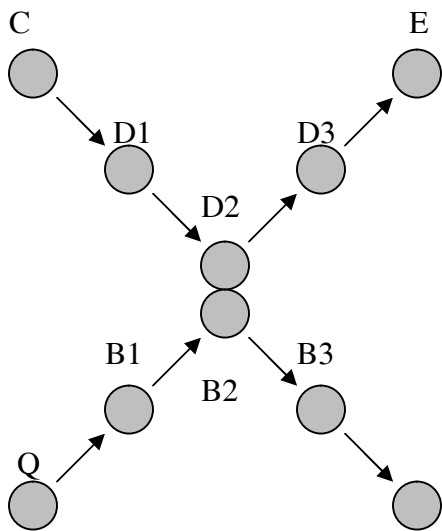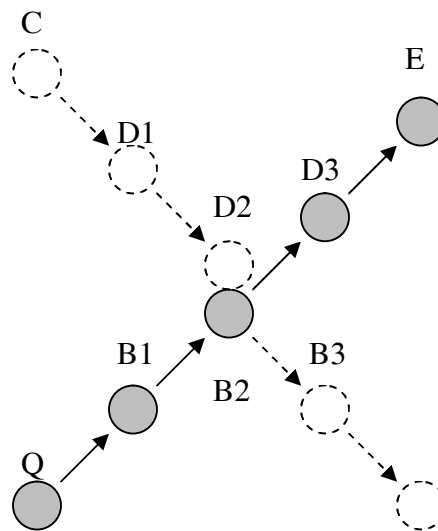[131] 'Middle pre-emption' seems an adequate name for this case.

Fig. 9: The closest ~*o*Q worlds to w0
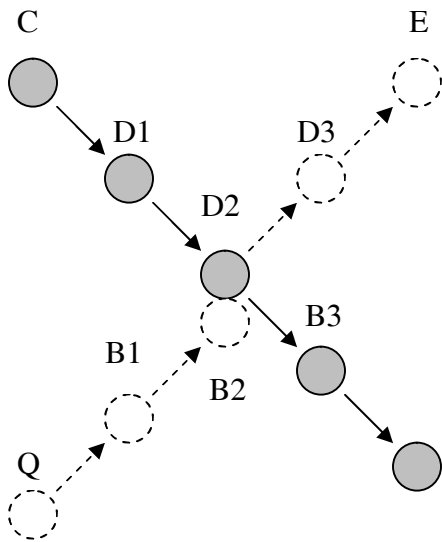


*Example*: Suppose Billy and Suzy are playing with rocks near a window. Suzy throws a rock (C), but aims away from the window. Billy throws a rock (Q), and aims for the window. As it happens, just after Suzy and Billy's rocks are in early flight (D1 & B1), both rocks collide in mid-air (D2 & B2), Suzy's rock rebounds towards the window in late flight (D3) shattering it (E), and Billy's rock flies away inertly (B3). In short, Suzy's throw (C) was a cause of the shattering (given it was her rock that shattered the window), but so is Billy's throw (Q) (given his rock enabled Suzy's rock to shatter the window (E) via the collision). However, if Suzy's throw (C) hadn't occurred then the shattering (E) still would have, given Billy's rock would have assumed a trajectory sufficient for the shattering (E). In addition, if Billy's throw (Q) hadn't occurred then the shattering (E) wouldn't have, given Suzy's rock wouldn't have rebounded off Billy's.

**2.5 Analysing the cases**

How does my analysis handle these four examples? We analyse Suzy's throw (C) as a cause of the shattering (E) in each case as follows (note it might be easier to keep just one of the above examples in mind for the following demonstration):

First, we note there is a process of Suzy's throw (C) to the shattering (E) in each case, as follows: Letting P be the chain of events of Suzy's throw (C), the rock in early flight (D1), mid-flight (D2), late flight (D3), and the shattering (E), and the background factors be Billy's throw/near throw (Q), the following counterfactuals are true:

1. If Suzy's throw (C), the rock in early flight (D1), mid-flight (D2), late flight (D3), and the shattering (E) had occurred and Billy's throw/near throw (Q) hadn't, then it would be the case that there is a corresponding chain of counterfactual dependence from Suzy's throw (C) to the shattering (E). As follows; if Suzy's throw (C) hadn't occurred then the rock in early flight (D1) wouldn't have, if its early flight (D1) hadn't occurred its mid-flight (D2) wouldn't have, if its mid-flight (D2) hadn't its late-flight (D3) wouldn't have, and if its late flight (D3) hadn't the shattering (E) wouldn't have. The counterfactual chain is complete; any event in any chain of dependence from Suzy's throw (C) to the shattering (E) is a counterpart to a part of the sum of events on the chain (C, D1, D2, D3, E).

2. If neither Suzy's throw (C), the rock in early flight (D1), mid-flight (D2), late flight (D3), nor the shattering (E) had occurred and Billy's throw (Q) had, then it

214

would be the case that there is a corresponding chain of proper counterfactual implication from Suzy's throw (C) to the shattering (E). As follows: If Suzy's throw (C) had occurred then the rock in early flight (D1) would have, if its early flight (D1) had occurred its mid-flight (D2) would have, if its mid-flight (D2) had its late-flight (D3) would have, if its late flight (D3) had the shattering (E) would have. The counterfactual chain is complete; any event in any chain of proper counterfactual implication from Suzy's throw (C) to the shattering (E) is a counterpart to a part of the sum of events on the chain (C, D1, D2, D3, E).

Thus, there is a process P from Suzy's throw (C) to the shattering (E). Is Suzy's throw (C) a cause of the shattering (E)? Suzy's throw (C) causes the shattering (E) in each case, as follows: There is a process P (C, D1, D2, D3, E) of Suzy's throw (C) to the shattering (E), and some background factors consisting of Billy's throw/near throw (Q), such that:

1. If the process P of Suzy's throw (C) to the shattering (E) had still occurred but Billy's throw/near throw (Q) hadn't, then if Suzy's throw (C) hadn't occurred the shattering (E) wouldn't have.

2. If the process P of Suzy's throw (C) to the shattering (E) hadn't occurred but Billy's throw/near throw (Q) had, then if Suzy's throw occurred the shattering would have.

Thus, Suzy's throw (C) causes the shattering in each case; the correct analysis. What about Billy's throw/near throw? In our cases of early and late pre-emption, there is no process from Billy's throw/near throw (re-termed event 'A') to the shattering (E). We achieve this result twice over, as follows:

1. No matter what chain (A, D, …, Dn, E) we pick out from Billy's throw/near throw (A) to the shattering (E), and no matter the background events (Q) we pick out, had such a chain occurred without such a Q, there would not be a chain of counterfactual dependence between actual events from the shattering (E) to Billy's throw/near (A), in which any event in any chain of counterfactual dependence of A to E features on the chain (A, D, …, Dn, E). The closest we get is counterfactual dependence of the shattering (E) on Billy's throw/near throw (A) in the case where Suzy's throw (C) is the background event (Q) we get rid of. But in such a case the chain of dependence would not be complete; given there would be events on some chain of dependence from the shattering (E) to Billy's throw/near throw (A) that don't feature on the actual chain (A, D, …, Dn, E); namely, events that invoke Billy's ball forcefully touching the window.

2. No matter what chain (A, D, …, Dn, E) we pick out from Billy's throw/near throw (A) to the shattering (E), and no matter the background events (Q) we pick out, had such a chain not occurred and the background events Q still had, there would not be a chain of (proper) counterfactual implication from his throw (A) to the shattering (E), in which any event in any chain of counterfactual implication

of A to E features on the chain (A, D, …, Dn, E). The closest we get is (proper) counterfactual implication from Billy's throw/near throw (A) to the shattering (E). But in such a case the chain would not be complete; given there would be events on some chain of (proper) counterfactual implication from Billy's throw/near throw (A) to the shattering (E) that do not feature on the actual chain (A, D, …, Dn, E); namely, events that invoke Billy's ball forcefully touching the window.

So we achieve twice over the result that there is no process from Billy's throw (A) to the shattering (E), and where there is no process there is no causation, and so we achieve the result that Billy's throw/near throw (A) was not a cause of the shattering (E); the correct analysis. Of course this isn't the case for the cases of double and middle pre-emption as far as Billy's throw is concerned. In the case of double pre-emption, the process from Billy's throw to the shattering is just like the process from Suzy's throw to the shattering, and so is analysed as a cause of the shattering just as Suzy's throw is. In the case of middle pre-emption, there is a process running from Billy's throw through to the collision with Suzy's ball and then from the flight of Suzy's ball to the shattering (consisting of events A, B1, B2 & D2, D3, E), and is analysed as a cause of shattering along with Suzy's throw.

So far the analysis of causation has met with success in cases of pre-emption which involve intrinsic causation. Cases of intrinsic causation provide the domain in which the intrinsic-ness thesis holds. The intrinsic-ness thesis holds that; for any collection of events S including an effect at t' together with all of its causes from some time t up until

t', any intrinsic duplicate of S with the same laws of nature has the same causal structure as $S^{132}$. Cases of extrinsic causation are cases in which the intrinsic-ness thesis fails. We shall now turn to cases of pre-emption that involve extrinsic causation.

## 2.6 Trumping pre-emption

Cases of trumping pre-emption are cases of pre-emption in which the cause's status 'trumps' the pre-empted cause's status from qualifying as a cause (see Schaffer [2000]).

Fig. 10: The actual world (w0)                Fig. 11: The closest ~C worlds to w0
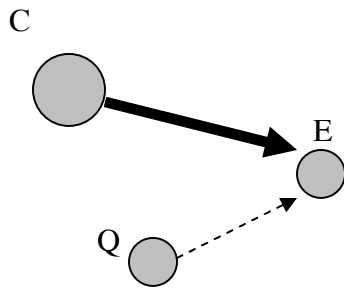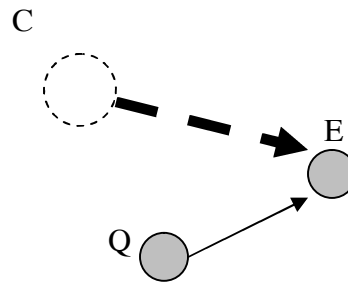


*Example*: Suppose it is a law of magic that 'the first spell cast on any given day matches the following enchantment that midnight'. Merlin casts the first spell that day to turn the prince into a frog (C), Morgana casts the same spell slightly later (Q), and at midnight, the prince turns into a frog (E). Merlin's spell, and not Morgana's, is the exclusive factor which causes the enchantment (it 'trumps' Morgana's spell), because the laws provide for

---

[132] See Hall, p.244 & p. 239, in Collins, Hall and Paul (ed.s) [2001].

the determination of the enchantment by Merlin's spell and not by Morgana's.[133] However, if Merlin's spell (C) hadn't occurred, the enchantment (E) still would have; given Morgana's spell (Q) would have sufficed for the enchantment (E).

## 2.7 Analysing the case

Merlin's spell (C) is analysed as a cause of the enchantment (E) as follows: there is a process P of Merlin's spell (C) to the enchantment (E) as follows: Letting P be the events of Merlin's spell (C) and the enchantment (E), and the background factors be Morgana's spell (Q), the following counterfactuals are true:

1. If Merlin's spell (C) and the enchantment (E) had occurred and Morgana's spell (Q) hadn't, then it would be the case that there is a corresponding chain of counterfactual dependence from Merlin's spell (C) to the enchantment (E); such that if Merlin's spell (C) hadn't occurred the enchantment (E) wouldn't have. The counterfactual chain is complete; any event in any chain of counterfactual dependence from Merlin's spell (C) to the enchantment (E) is a counterpart to a part of the sum of events on the chain (C, E).

2. If neither Merlin's spell (C) nor the enchantment (E) had occurred but Morgana's spell (Q) had, then it would be the case that there is a corresponding

---

[133] We suppose that Merlin's spell causes the enchantment directly at a spatio-temporal distance by appeal to no intermediary events.

chain of (proper) counterfactual implication from Merlin's spell (C) to the enchantment (E); such that if Merlin's spell (C) had occurred the enchantment (E) would have. We note that the counterfactual chain is complete; any event in any chain of (proper) counterfactual implication from Merlin's spell (C) to the enchantment (E) is a counterpart to a part of the sum of events on the chain (C, E).

Why is the second condition satisfied? According to our Lewisian-style counterfactual semantics, we must appeal to the closest worlds in which Morgana's spell (Q) occurs without Merlin's spell (C) or the enchantment (E). What do these worlds look like? If we are to take our Lewisian similarity relation[134] seriously, they must be worlds in which the laws are locally 'violated' - affording us a local law-like 'divergence miracle' to use Lewis' phrase. So the relevant law now becomes '(for all spells *apart* from Morgana's spell) the first spell cast on any given day matches the following enchantment that midnight'. We then test whether Merlin's spell counterfactually implies the enchantment in such worlds, and it turns out it does, given no law-like exceptions apply to Merlin's spell.

Thus there is a process P of Merlin's spell (C) to the enchantment (E). Does Merlin's spell (C) cause the enchantment (E)? Recalling our analysis of processes, we note that Merlin's spell (C) causes the enchantment (E), as follows: there is a process P (C, E)

---

[134] See Lewis [1979].

from Merlin's spell (C) to the enchantment (E) and some background factors of Morgana's spell (Q) such that:

1. If the process P (i.e. Merlin's spell (C) and the enchantment (E)) had occurred but Morgana's spell (Q) hadn't, then it would be the case that if Merlin's spell hadn't occurred then the enchantment (E) wouldn't have.

2. If the process P (i.e. Merlin's spell (C) and the enchantment (E)) hadn't occurred but Morgana's spell (Q) had, then it would be the case that if Merlin's spell had occurred then the enchantment (E) would have.

Thus, Merlin's spell (C) causes the enchantment (E); the correct analysis. How do we analyse Morgana's spell as a non-cause of the enchantment? We observe that for the enchantment (E) to depend on Morgana's spell (now A), and thus for our first condition of processes to go through, we must 'get rid' of Merlin's spell (now Q) first, as follows:

1. If Morgana's spell (A) and the enchantment (E) had still occurred but Merlin's spell (Q) hadn't, then it would be the case that there is a corresponding chain of counterfactual dependence from Morgana's spell (A) to the enchantment (E); such that if Morgana's spell (A) hadn't occurred the enchantment (E) wouldn't have. The counterfactual chain is complete; any event in any chain of counterfactual dependence from Morgana's spell (A) to the shattering (E) is a counterpart to a part of the sum of events on the chain (A, E).

The problem now is that the second condition for processes fails, as follows:

> 2. If neither Morgana's spell (A) nor the enchantment (E) had occurred but Merlin's spell still had, then it would *not* be the case that there is a corresponding chain of proper counterfactual implication from Morgana's spell (A) to the enchantment (E); given if Morgana's spell (A) had occurred the enchantment (E) wouldn't have.

Why is this true? According to our Lewisian-style counterfactual semantics, we must appeal to the closest worlds in which neither Morgana's spell nor the enchantment (E) occurs but Merlin's spell still does. What do these worlds look like? If we are to take our Lewisian similarity relation seriously, they must be worlds in which the laws are locally violated - affording us a local law-like 'divergence miracle' to use Lewis' phrase. So our laws become '(for all spells *apart* from Merlin's spell) the first spell cast on any given day matches the following enchantment that midnight'.[135] We then test whether Morgana's spell counterfactually implies the enchantment in such worlds, and it turns out

---

[135] Rules for how divergence laws operate are wanting. Jonathan Schaffer has suggested (in correspondence) that it isn't obvious why the relevant law shouldn't alternatively become 'the first (except for Merlin's) spell cast on any given day matches the following enchantment that midnight'. In Schaffer's scenario, Morgana's spell would troublesomely become active; given it would be the first – except for Merlin's - spell that day. We avoid this issue simply by specifying a preference as to what the rules for divergence miracles are, in the ways I suggest. Following a Lewisian spirit, I maintain this preference is a case of 'spoils to the victor', for an otherwise successful theory of causation (see Lewis p. 194 [1986a]).

it doesn't, given Morgana's spell would not be the first spell that day (and so isn't

subsumed by our law of magic), and because the first spell that day (Merlin's spell) still

fails to determine the enchantment given the locally divergent laws. Thus, there is no

process (and thus no causation) from Morgana's spell to the enchantment (E); the correct

analysis. We move on to one more case of pre-emption before some closing remarks.


## 2.8 Pre-emptive prevention


Cases of pre-emptive prevention are cases in which an event prevents another, but had

the former not prevented the latter, another event would have (see Collins [2000]).


Fig. 12: The actual world (w0)                Fig. 13: The closest ~C worlds to w0
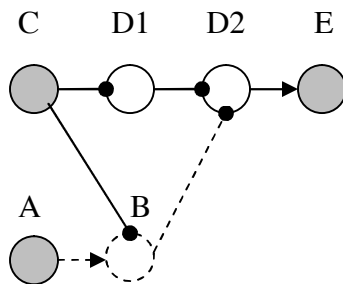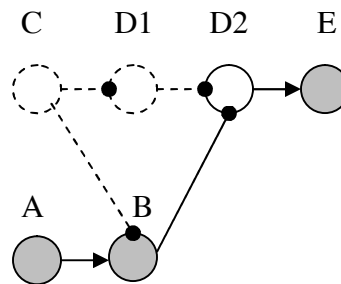


*Example*: Suppose Suzy is standing in front of a window, and Billy is standing a short

distance away from Suzy and the window (but not in the ball's line of fire). A ball is

thrown, and had its flight been uninterrupted (without Suzy or Billy ready to intervene), it

would have shattered the window. Suppose Suzy catches the ball (C), but if Suzy hadn't

Billy would have jumped in and caught the ball (B) in turn preventing the shattering. All

things being equal, we say Suzy's catch (C) (and not Billy's) prevents the shattering of the window, or alternatively, causes the window's preservation (E).

It is worth contrasting this example against a similar one in which Suzy's catch doesn't prevent the shattering:

*Example*: Suppose Suzy is standing in front of a massive wall, and the wall is in front of a window. A ball is thrown, and had its flight been uninterrupted (without Suzy or the wall ready to intervene), it would have shattered the window. Suppose Suzy catches the ball (C), but if Suzy hadn't, the ball would have rebounded off the wall in turn preventing the shattering. All things being equal, we say Suzy's catch (C) *doesn't* prevent the shattering of the window, or alternatively, *doesn't* cause the window's preservation (E).

## 2.9 Analysing the case

Why does Suzy's catch prevent the shattering in the first case but not in the second? Suzy's catch prevents the shattering in the first case, partially because the pre-empted preventer (Billy's near catch) *was not already in place* to do any preventing (Billy was nowhere near the ball throughout the scenario), and partially because Suzy eliminated a *relatively major threat* to the window (the flying ball). In contrast, Suzy's catch didn't prevent the shattering in the second case, partially because the pre-empted preventer (the presence of the wall) *was already in place* to do some preventing (the wall doesn't need to do anything but stand there to prevent the ball from shattering the window), and

partially because Suzy did *not* eliminate a *relatively major threat* to the window (the window was not really threatened by the ball; far more likely it would be that greater threats from potentially aggressive neighbours came from elsewhere, no matter how remote). Happily, my analyses confirm these conclusions. The first example first:

Recalling our analysis of processes, we note that there is a process P of Suzy's catch (C) to the window's preservation (E) as follows: Letting P be the chain of occurrences of Suzy's catch (C) and the non-shattering (E), and the background events be Billy's background presence (Q) in which he stands his distance, the following counterfactuals are true:

1. If Suzy's catch (C) and the window's preservation (E) had occurred but Billy's background presence (Q) hadn't, then it would be the case that there is a chain of counterfactual dependence from Suzy's catch (C) to the window's preservation (E); such that if Suzy's catch (C) hadn't occurred, then the window's preservation (E) wouldn't have. The counterfactual chain is complete; any event in any chain of counterfactual dependence from Suzy's catch (C) to the window's preservation (E) is a counterpart to a part of the sum of events on the chain (C, E).

2. If neither Suzy's catch (C) nor the window's preservation (E) had occurred but Billy's presence (Q) had, then it would be the case that there is a chain of proper counterfactual implication from Suzy's catch (C) to the window's preservation (E); such that if Suzy's catch (C) occurred, then the window's preservation (E)

225

would have. The counterfactual chain is complete; any event in any chain of proper counterfactual implication from Suzy's catch (C) to the window's preservation (E) is a counterpart to a part of the sum of events on the chain (C, E).

Why is the second condition satisfied? According to our Lewisian-style counterfactual semantics, we must appeal to the closest worlds in which Billy's background presence (Q) still occurs without Suzy's catch (C) or the window's preservation (E). What do these worlds look like? If we are to take the Lewisian similarity relation seriously, they must be worlds in which the ball that Suzy now fails to catch shatters the window, given we have 'held fixed' the events in which Billy inertly waits in the background without interfering with the ball's trajectory (like he does in the actual world). We then test whether Suzy's catch (C) properly counterfactually implies the window's preservation (E) in such worlds, and it turns out it does, given that Billy still waits inertly in the background.

Thus there is a process P of Suzy's catch (C) to the window's preservation (E); the correct analysis. Does Suzy's catch (C) cause the window's preservation (E)? Recalling our analysis of causation, we note that Suzy's catch (C) causes the window's preservation (E), as follows: There is a process P (C, E) of Suzy's catch (C) to the window's preservation (E) and some background events of Billy's background presence (Q) such that:

1. If the process (P) (Suzy's catch (C) and the window's preservation (E)) had occurred but Billy's background presence (Q) hadn't, then it would be the case

that if Suzy's catch (C) hadn't occurred then the window's preservation (E) wouldn't have.

2. If the process (P) (Suzy's catch (C) and the window's preservation (E)) had not occurred but Billy's background presence (Q) had, then it would be the case that if Suzy's catch (C) occurred then the window's preservation (E) would have.

Thus Suzy's catch (C) causes the window's preservation (E) in our first example; the correct analysis. How do we analyse her catch as a non-cause of the non-shattering in our second example? We first observe that for the window's preservation (E) to depend on Suzy's catch (C), and thus for our first condition of processes to go through, we must 'get rid' of the wall's presence (Q), as follows:

1. If Suzy's catch (C) and the window's preservation (E) had occurred but the wall's presence (Q) hadn't, then it would be the case that there is a chain of counterfactual dependence from Suzy's catch (C) to the window's preservation (E); such that if Suzy's catch (C) hadn't occurred, then the window's preservation (E) wouldn't have. The counterfactual chain is complete; any event in any chain of counterfactual dependence from Suzy's catch (C) to the window's preservation (E) is a counterpart to a part of the sum of events on the chain (C, E).

The problem now is that the second condition for processes fails, as follows:

2. If neither Suzy's catch (C) nor the window's preservation (E) occurred but the wall's presence (Q) had, then it would *not* be the case that there is a chain of proper counterfactual implication from Suzy's catch (C) to the window's preservation (E); given if Suzy's catch (C) occurred, then the window's preservation (E) wouldn't have.

Why is this true? According to our Lewisian-style counterfactual semantics, we must appeal to the closest worlds in which the wall's presence (Q) still occurs without Suzy's catch (C) or the window's preservation (E). What do these worlds look like? If we are to take the Lewisian similarity relation seriously, they must be worlds in which the ball Suzy now fails to catch bounces off the wall, but in which the shattering occurs for far more different reasons; far closer is a non-catch, wall, and shattering world in which a far greater alternative threat to the window shatters it (no matter how remote), than a non-catch, wall, and shattering world in which the uncaught ball slices through or around a massive wall in a relatively implausible manner. We then test whether Suzy's catch (C) properly counterfactually implies the window's preservation (E) in such worlds, and it turns it doesn't, given the greater alternative threat would still shatter it. Thus, there is no process or causation from Suzy's catch (C) to the window's preservation (E) in our second case; the correct analysis.

## 3. Concluding remarks

Previous counterfactual analyses of deterministic causation have been unsuccessful in handling all the pre-emption cases. In this chapter, I've presented a new counterfactual analysis of deterministic causation that demonstrably succeeds where the others have failed. The analysis I offer suggests that we should mind our Ps and Qs; causation seems to be a matter of processes (Ps) and background factors (Qs) that exhibit the right pattern of would-be counterfactuals. Amongst these would-be counterfactuals, counterfactual dependence and implication play an equal role, an equal role that points to where past strategies have gone so wrong. Past strategies have focused solely on adding 'bells and whistles' to conditionals of would-be counterfactual *dependence* in order to analyse causation, where this is really only one side of the coin. The other side is to add 'bells and whistles' to conditionals of would-be counterfactual *implication*, and it was with this moral in mind that I suggested the ongoing debate should be steered.[136]

**Further Discussion**

It should be clear how my new theory takes hold of the lessons and morals I suggested earlier. As follows:

---

[136] Might there be counter-examples to the theory I offer above? Schaffer has suggested two types of worries to me. The first worry is that maybe there are structures of events that *mimic* a causal process, without actually being one. The second worry is to do with the way I deal with cases of *trumping*. Might causes of Merlin's spell which are also causes of the enchantment be analysed incorrectly as non-causes of the enchantment? Unfortunately I haven't been able to conceive of examples of these forms which in turn create trouble for my analysis.

Firstly, I said that a successful analysis should be easily revisable to accommodate a non-transitive and transitive relation. My analysis does this, for where my analysis analyses a non-transitive causal relation, it is easily revisable to provide for a transitive causal relation; to do so we simply take chains of my non-transitive causal relation to analyse a transitive causal relation.

Secondly, I said that a successful analysis should be easily revisable to i) accommodate both causation between fine and coarse grained events, and ii) should allow omissions as the relata. i) is easily dealt with; my analysis has not said whether we take fine or coarse grained events as the relata, and so we may impose whatever conditions we wish with respect to the relata's granularity. ii) requires more discussion. For the sake of simplicity, I only considered causation between events when setting out my theory above, but causation of, by, and via omission is also possible (see, Lewis [2001b], Schaffer [2000], and Beebee [2004] for discussion, and my section on preemptive prevention). For my analysis of causation to capture causation that includes omissions, I require only a minor revision to my analysis of processes; the revision is that the chain (C, D,…, Dn, E) is a chain (some of the occurrences[137] of which are distinct) which includes the actual occurrences of C & E, and if anything else, actual events. This revision thus takes a process of C to E to consist of C, E (whether they be omissions or events) together with all the events of a causal chain from C to E (if there are any). This revision suffices to

---

[137] An occurrence is an event or an omission. Events are intrinsic properties local to space and time (see Lewis [1986b]), omissions are the absence of intrinsic properties local to space and time.

deal with ii), albeit at the expense of having a gerrymandered analysis of processes- so be it.

Thirdly, I said that a successful analysis, if it appeals to processes in its analysis, should provide clear truth conditions for what indeed constitutes a process. I provided such an analysis earlier.

Fourthly, I said that a successful analysis may do well to take advantage of 'getting rid of' and 'holding fixed' strategies. These strategies are employed throughout my analyses of causation and processes. Witness my analysis of causation; condition 1 uses a strategy of 'getting rid' of troublesome background events, 'holding fixed' the process of cause to effect, and then testing for counterfactual dependence of the effect on the cause. Condition 2 uses a complementary strategy of 'getting rid' of the process of cause to effect, 'holding fixed' the troublesome background events, and then testing for counterfactual implication from cause to effect.

Fifthly, I said that a successful analysis might do well to employ a strategy of 'embedded counterfactuals'. This strategy is employed throughout my analyses of causation and processes - each of the two conditions for causation/processes employs embedded counterfactuals to a similar effect as the would be dependence analysis.

Sixthly, I said that a successful analysis might have an element that we have not yet uncovered, the intuition being that this missing element has something to do with

counterfactuals. The element, as I suggested above in the analysis, is to take hold of conditionals of counterfactual implication as much as counterfactual dependence.

I now conclude the thesis with some general remarks and observations.

**4.2 Conclusion**

This thesis' major project has been the investigation of whether anything like a counterfactual analysis of causation might succeed in meeting Lewis' requirements. Lewis' requirements were that a counterfactual analysis of causation must, along with analysing basic cases of causation correctly:

- Deal with the problems of effects and epiphenomena
- Deal with the problem of pre-emption, and
- Be grounded in the intuition that causation has something to do with regularities and counterfactuals, without piling on the epicycles.

We have had (for the sake of argument) to proceed as if counterfactuals came equipped with a theory for the causal asymmetry that in turn satisfactorily dealt with the problems of effects and epiphenomena: insofar as we have not given an answer to this problem, our results are incomplete. Finding an answer to this problem might not be an easy project for the counterfactual strategy; as Noordhof observes in 'Prospects for a counterfactual

analysis of causation',[138] the issue of how counterfactual strategies can cope with analysing causal asymmetries is one of the most pressing ones for the counterfactual approach to causation to date. I have had to leave this issue aside.

Nevertheless, I think considerable progress has been made in this thesis with respect to dealing with the problem of pre-emption, and towards generating analyses of causation that can handle these cases in a way that doesn't pile up the epicycles in a seemingly ad hoc way. After much investigation into the pre-emption cases, I suggested some key morals for what a successful analysis of causation would look like. In chapter 4, I presented a new counterfactual analysis of causation that could deal with all the deterministic pre-emption cases in the literature (and then some!) based on these morals. The results for just some of the analyses of causation considered are catalogued as follows:

---

[138] See Dowe, P. & Noordhof, P. eds. (2007), 'Cause and Chance; Causation in an indeterministic world.'

| | CFA | FGCFA | CFCA | QDA | RQDA | WBDA | RWBDA | ICA | HFIA | HFA | FGHFA | ISA | PQA |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CG-EP | x | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| FG-EP | x | x | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| NL-EP | x | x | x | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| MF-EP | x | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | x | ✓ | ✓ | ✓ | ✓ | ✓ |
| OG-EP | x | x | ✓ | ✓ | ✓ | ✓ | ✓ | x | x | ✓ | ✓ | ✓ | ✓ |
| S-EP | x | x | ✓ | ✓ | ✓ | x | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| CG-LP | x | ✓ | x | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | x | ✓ | ✓ | ✓ |
| FG-LP | x | x | x | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | x | ✓ | ✓ | ✓ |
| MF-LP | x | x | x | ✓ | ✓ | ✓ | ✓ | x | x | x | ✓ | ✓ | ✓ |
| OG-LP | x | x | x | x | ✓ | x | ✓ | x | x | x | ✓ | ✓ | ✓ |
| S-LP | x | x | x | x | ✓ | ✓ | ✓ | ✓ | ✓ | x | ✓ | ✓ | ✓ |
| B-MP | x | x | x | ✓ | ✓ | x | ✓ | x | ✓ | ✓ | ✓ | ✓ | ✓ |
| S-MP | x | x | x | ✓ | ✓ | x | ✓ | x | x | x | ✓ | ✓ | ✓ |
| B-TP | x | x | x | x | x | x | x | ✓ | ✓ | x | x | x | ✓ |
| S-TP | x | x | x | x | x | x | x | x | x | x | x | x | ✓ |
| B-PP | x | x | x | ✓ | ✓ | ✓ | ✓ | x | x | ✓ | ✓ | ✓ | ✓ |
| S-PP | x | x | x | x | x | x | x | x | x | ✓ | ✓ | ✓ | ✓ |
| B-DP | x | ✓ | x | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | x | x | ✓ | ✓ |
| S-DP | x | x | x | ✓ | ✓ | ✓ | ✓ | x | x | x | x | ✓ | ✓ |
| DPV | ✓ | ✓ | ✓ | x | x | x | x | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |

(CFA = (Hume's) counterfactual analysis, FGCFA = The fine-grain counterfactual analysis, CFCA = (Lewis') counterfactual chains analysis, QDA = (Lewis') quasi-dependence analysis, RDQA = the revised quasi-dependence analysis, WBDA = (Collins') would-be dependence analysis, RWBDA = the revised would-be dependence analysis, ICA = (Lewis') influence chains analysis, HFIA = the holding fixed influence analysis, HFA = the holding fixed analysis, FGHFA = the fine-grain holding fixed analysis, ISA = the isostructural analysis (see appendix 5.2), PQA = my own 'Ps and Qs' analysis. Suffixes EP, LP, MP, TP, PP, DP, stand for early pre-emption, late pre-emption, middle pre-emption, trumping pre-emption, pre-emptive prevention, and double pre-emption respectively. Prefixes CG, FG, NL, MF, OG, B, S stand for coarse-grained, fine-grained, non-local, modally fragile, over-generative, basic, and strengthened respectively. DPV stands for double prevention).

As we can see, only my analysis deals with all the pre-emption cases correctly. Is it, then, the correct analysis of causation? Providing a correct theory for the causal asymmetry can be provided for it, I suggest so. But failing this, the early successes of my analysis confirm the utility of the morals and lessons I provided earlier. Indeed, if the counterfactual strategy is to be attempted, it seems that strategies that appeal to processes, that appeal to a 'holding fixed' and 'getting rid of' troublesome background pre-empted causes, and that appeal to conditionals of would-be counterfactual *implication* along with conditionals of would-be counterfactual *dependence*, prove to be an exceptionally profitable approach to take.

## 5. Postscript: Towards a theory of Causation and Background factors

What is it for a thing to be a cause, as opposed to a mere causal condition or background factor for an effect? We tied this issue to the issue of causal *selection* in the introduction, an issue I have up until now ignored. Is there no serious distinction to be had (see Lewis [1986a])? Is it grounded in the activities of agents (see Hart and Honour [1959])? Is it determined contrastively (see Schaffer [2005])? Or something else? In this short postscript, I would like to tentatively outline a theory which ties the distinction to the latter - namely, I would like to sketch a view that says that backtracking counterfactuals are intimately linked to the distinction between causes and background conditions.

I introduced my analysis of causation in chapter 4, and perhaps my analysis is better described as an analysis of causal conditions, than of full-blown causation. This is because my theory affirms the well worn (and perhaps incorrect) Lewisian theory that counterfactual dependence between distinct actual events is sufficient for causation. I strike a match, it burns - had I not struck the match it wouldn't have burnt, but had there not been oxygen the match wouldn't have burnt either, thus the striking and the presence of oxygen were both causes, or - perhaps more accurately - causal conditions of the burning. Whether we agree on the fact that the oxygen and the flame share the status of causes or just causal conditions or not, what we might agree on is that the striking and the oxygen have different causal statuses; the striking was the main foreground cause/causal condition of the flame, the presence of oxygen a mere background cause/causal condition. It's a very general outline of a theory for this distinction, in terms of

backtracking counterfactuals, that I now turn. I think an investigation into such an outline is important as, so far, no one has attempted to give a counterfactual analysis for this distinction, where such an analysis would obviously tie in well with the spirit of a counterfactual analysis of causation.

Here is the main thesis: All things being equal, what distinguishes the causal status of the striking from the causal status of the presence of oxygen in the above example are some backtracking counterfactuals – it's true that if the match hadn't lit, then it wouldn't have been struck, but it's not true that if the match hadn't lit, there wouldn't have been oxygen. As we might expect, backtracking counterfactuals (which are usually counterfactuals in which the antecedent is temporally after or caused by the consequent) must be given a very different analysis to the Lewisian forward-tracking counterfactuals we have been visiting. Unfortunately, no theory for backtrackers (to my knowledge) is available in the literature. This shouldn't deter us from using them however; so long as the backtrackers we appeal to form a safe part of the cases which a successful analysis of backtrackers should analyse, nothing mysterious has been said, and we are treading on safe ground.[139]

---

[139] Menzies (p.149-150, [1989]) observes that backtracking reasoning, unguided by clear cut semantics, can lead to opposing conclusions in some cases. I agree with him; had the shattering Suzy caused not occurred, would there or wouldn't there have been the big bang? I'd say the big bang still would have occurred, but others may disagree. But if they disagree then they may well be using a different type of backtracker. In any case, the following demonstrations are merely meant to serve as an illustration of the right directions the debate should take with regard to the use of backtrackers, where unequivocal answers in all cases will come from unequivocal semantics for the required sort of backtracking conditional; something we will just have to wait for.

Indeed, if the use of intuitive backtrackers doesn't yield intuitive, clean cut results with regard to the concept we are analysing, then that would be a very good reason to criticise the theory that appeals to them for want of a decisive semantics to make the call. It's a theory for the distinction by appeal to backtrackers which I wish to outline.

(If there had been a problem here with using backtrackers, then it would've been the case that they never made good sense in the first place. Perhaps, it might be complained, if backtrackers had now made good sense, then Lewis would have written more about them. So I would apologise now if they seem unclear after their forthcoming use. In an ideal world, if I had still appealed to backtrackers I would have done so with a prior appeal to a worked out semantics from the literature. But in any case if I *had* to convince you that backtrackers made good sense, I wouldn't have previously foreseen a way to do so without a semantical story. Or perhaps I would have; the last five sentences were backtracking counterfactuals, so if you understood my apology I now retract it; backtrackers do in fact make good sense to you.)

Here are some theories (still restricted to the deterministic case) that appeal to backtrackers, where the notion of similarity is to be given a yet undiscovered and non-Lewisian analysis:

C is a foreground causal condition of E iff

- C causes E, & if E hadn't occurred then C wouldn't have (i.e. the most similar ~E worlds are ~C worlds).

C is a background causal condition of E iff

- C causes E, & if E hadn't occurred then C still would have (i.e. the most similar ~E worlds are C worlds).

These theories come in handy for promoting and demoting causal conditions to a deserved causal status.[140] Here are 6 examples for good measure; the first two involve causation by omission, the next four involve causation by events:

---

[140] This isn't the only way causes can differ in causal status: sometimes they can differ in causal status not only as a discrete (all or nothing) affair provided for by the distinction of back/foreground conditions, but sometimes they can differ in causal status as a continuous (matter of degree) affair provided for by the distinction of being a relatively good/bad causal explainer. I have a proto-theory for causal explanation, but it derives its strength from the notion of closeness that our backtrackers use. As I don't want to put too much stress on our under-analysed notion of closeness, I only footnote the theory here: C causally explains E more than D does iff either C causes E and D doesn't, or C causes E and E depends more on C than D. Furthermore, E depends more on C than D iff the closest (~E,~C) worlds are closer than the closest (~E,~D) worlds. Provided we appeal to our backtracking notion of closeness, we can already see that foreground causes will always be better causal explainers than background causes, for where C is a foreground cause and D is a background cause, a ~E, ~C world will be closer than any ~E,~D world given ~E worlds will automatically be ~C worlds and not ~D worlds.

*Example 1*: Suppose Billy and Suzy plan to meet up for lunch, but that Billy doesn't turn up; making Suzy sad. Suppose also that Bill Gates didn't turn up to lunch with a $10 million cheque for Suzy and that he certainly wasn't in a position to. Two Lewisian counterfactuals are true:

- If Billy's turning up to lunch occurred, Suzy would not have been sad
- If Bill Gates had turned up with a $10 million cheque for Suzy, Suzy would not have been sad

The forward-trackers confirm two causal instances; the omission of Billy's turning up to lunch is a causal condition of Suzy's being sad, and so is the omission of Bill Gates and his exorbitant cheque; but backtrackers provide a difference for causal status. Billy's not turning up for lunch is a main foreground cause of Suzy's being sad; had Suzy not been sad, Billy would have turned up (i.e. the most similar worlds in which Suzy isn't sad are worlds in which Billy turns up). Bill Gates' not turning up for lunch with an exorbitant cheque is a background cause of Suzy's being sad; had Suzy not been sad, Billy would have turned up and not Bill Gates (i.e. the most similar worlds in which Suzy isn't sad are worlds in which Bill Gates still doesn't turn up).

*Example 2*: Suppose Suzy regularly waters her neighbour's plants. Suppose she stops watering them just because she forgets (something she isn't otherwise likely to do), making them die. Suppose the Queen didn't water the plants either and she certainly wasn't expected to, and that she certainly wasn't in a position to. We suppose the most

240

realistic way in which the plants would have continued to survive is if Suzy hadn't forgotten to water the plants. Two Lewisian counterfactuals are true:

- If Suzy had watered the plants, then they wouldn't have died
- If the Queen had watered the plants, then they wouldn't have died

The forward-trackers confirm two causal instances; the omission of Suzy's watering is a causal condition of the plants dying, and so is the omission of the Queen turning up with a watering can; but backtrackers provide a difference for causal status. Suzy's not watering the plants is a main foreground cause of the plants dying; had the plants not died, Suzy would have watered them (i.e. the most similar worlds in which the plants hadn't died are worlds in which Suzy waters them). The Queen's not turning up with a watering can is a background cause of the plants dying; had the plants not died, Suzy would have watered them and not the Queen (i.e. the most similar worlds in which the plants don't die are still worlds in which the Queen doesn't turn up).

*Example 3*: Suppose Billy smokes 40 a day and this results in him getting lung cancer. Two counterfactuals are true:

- If Billy hadn't smoked 40 a day, he wouldn't have got lung cancer
- If Billy hadn't possessed lungs, he wouldn't have got lung cancer

The forward-trackers confirm two causal instances; Billy's smoking 40 a day caused him to get lung cancer, and so did Billy's possession of lungs; but backtrackers provide a difference for causal status. Billy's smoking is a main cause of his getting lung cancer; had Billy not got lung cancer he wouldn't have smoked (i.e. the most similar worlds in which Billy didn't have lung cancer are worlds where he didn't smoke 40 a day). Billy having lungs is a background causal condition to his getting cancer; had he not had lung cancer he still would have had lungs (i.e. the most similar worlds in which Billy hadn't got lung cancer are worlds in which Billy still has lungs).

*Example 4*: Suppose that there are heavy rains in winter, which delay the annual forest fires (caused by the scorching sun in June) from May until June. Two Lewisian counterfactuals are true:

- If the sun hadn't scorched the forest in June, there wouldn't have been the June forest fires
- If there hadn't been heavy winter rains, then there wouldn't have been the June forest fires (they would have happened in May instead)

The forward-trackers confirm two causal instances; the scorching June sun caused the June fires, but so did the heavy rains the previous winter; but backtrackers provide a difference for causal status. The scorching June sun is a main foreground cause of the forest fires; had there been no June forest fires there wouldn't have been a scorching June sun (i.e. the most similar worlds in which there are no June forest fires are worlds in

which there is no scorching June sun). The heavy winter rain was only a background causal condition to the June forest fires; had there been no June forest fires there still would have been heavy winter rains (i.e. the most similar worlds in which there are no June forest fires are worlds in which there are previous heavy winter rains).

*Example 5*: Suppose the 8 ball is at rest in the middle of the pool table, and Suzy shoots at it with the white causing it to move across the table. Two Lewisian counterfactuals are true:

- If Suzy had not shot at the 8 ball, there would be no moving 8 ball after Suzy's shot
- If the 8 ball had not existed before Suzy made her shot, there would be no moving 8 ball after Suzy's shot

The forward-trackers confirm two causal instances; Suzy's shot was a causal condition for the moving 8 ball, but so was the existence of the 8 ball before Suzy's shot; but backtrackers provide a difference for causal status. Suzy's shot is a main foreground cause of the moving 8 ball; had there been no moving 8 ball Suzy wouldn't have made her shot (i.e. the most similar worlds in which there is no moving 8 ball are worlds in which Suzy doesn't make her shot). The existence of the 8 ball before Suzy's shot is a mere background condition of the moving 8 ball (i.e. the most similar worlds in which the 8 ball hadn't moved are still worlds in which the 8 ball existed before Suzy's shot).

Not all causes can be sharply divided into background and foreground causes. Sometimes, it is indeterminate whether a causal condition is a foreground or background causal condition, even in worlds that are deterministic. Here is one such example:

*Example 6*: Suppose a 2 dimensional, spatio-temporally discrete world in which the laws about small bits of matter are governed much like that of the rules of John Conway's 'Game of Life': the world is made of a massive grid of squares in which the following deterministic law holds; if 4 squares surrounding an empty square are each occupied by a blob, then at the next time the empty square will be occupied by a blob, and the surrounding blobs will disappear (suppose the massive grid is large and varied enough to confirm these laws). Now consider the following states of affairs local to one 3 by 3 grid.
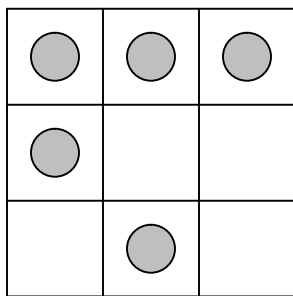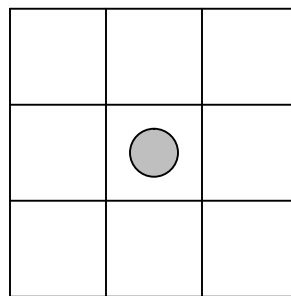
Time $t$                    Time $t$ +1

The following forward-trackers are true: Had any of 10 combinations of two blobs, or any of 9 combinations of 3 blobs, or any of 5 combinations of 4 blobs, or all 5 blobs at $t$ not occurred, then the blob at $t$ +1 would not have occurred. This makes for 25 different

causal relationships between different combinations of blobs at $t$ and the blob at $t$ +1. Which blobs at $t$ are the main causes and which ones the background conditions? The following backtracker is true: had there not been a blob at $t$ +1, then there wouldn't have been 4 blobs at $t$; i.e. the most similar possible worlds in which there wasn't a blob at $t$ +1 are worlds in which there weren't 4 blobs at $t$. What do these worlds look like? The most similar worlds in which there aren't 4 blobs at $t$ are worlds in which there are only 3 blobs at $t$, and there are 9 different worlds of this sort. Thus, if the event of there being 4 blobs at $t$ is a main cause of there being a blob at $t$ +1, what are the background causes? Are, say, the two top blobs at the top left hand side of the grid at time $t$ a background cause or a foreground cause? It's indeterminate - i.e. it's true that it is either a background condition or a main cause but it's just not determinate which. Why? Had the blob at $t$ +1 not occurred, then it would be indeterminate whether those two particular blobs would still have occurred (or in possible worlds jargon; some of the most similar worlds in which there wasn't a blob at $t$ +1 are worlds in which there aren't those two blobs at t, and some of them are). To generalise, we then have a third theory about the statuses of causal conditions, as follows:

- C is an indeterminate foreground/background causal condition of E iff C causes E, and if E had not occurred it would be indeterminate whether C still would have (i.e. some of the most similar ~E worlds are C worlds and some of them aren't).[141]

---

[141] The indeterminacy here is some sort of indeterminacy- I leave it open as to what type of indeterminacy it exactly is.

**Concluding discussion:**

I think the backtrackers I provide with respect to all these cases are pretty clear and unequivocal on their results. What conclusions can we now draw? For starters, I think that these early results suffice to justify my major goal; that backtracking counterfactuals seem to be intimately linked with the distinction between causes and background conditions.

Something also worth discussing is that the backtracking analyses I have offered may very well break down when we visit cases of *indeterministic* causation upon the appropriate analysis of similarity. Suppose there was only a 10% chance that my striking a match would result in a burning (and that I strike a match, and it burns); you might think it is true that had the match not lit, I still would have struck but that my striking would have simply failed to result in a burning (we might think this because, following the analysis of backtrackers I have offered- worlds in which I strike a match and the match does not burn are closer than worlds in which I don't strike a match and it doesn't burn). This is not a problem: here's a simple revision to fit indeterministic cases such as this one; we say that C is a foreground cause of E iff C causes E, and if E *couldn't* have occurred then C wouldn't have (i.e. the closest worlds in which E is nomologically impossible are ~C worlds). Upon this analysis (and to return to our indeterministic case), if the lighting of the match couldn't have occurred the striking wouldn't have, because a striking would have allowed for the lawful possibility of the lighting, and so the striking

is analysed as a foreground cause of the lighting in the indeterministic case- the correct analysis.

Obviously, the project of using backtrackers to provide us with the distinction between foreground and background causes is in need of extensive development. For starters, a precise semantics for backtracking counterfactuals must be given. The outstanding issue in providing such a semantics is to find an answer to the question of what plays the role of *closeness* in such conditionals (that is, if we are to analyse backtrackers in a similar way to how Lewis analyses forward-trackers). I have tentatively suggested a non Lewisian notion of similarity, but perhaps a notion of naturalness, normalcy, or familiarity will do - I won't make a choice here; it's yet unclear. What is suggested is that intuitive backtrackers seem to yield the correct results in paradigm cases of causation where there is a distinction to be made between main causes and background conditions. But what is also interesting is that this suggestion seems to go both ways; it also seems that cases in which there is a distinction to be made between main causes and background conditions suggest something about what a correct theory of backtracking counterfactuals should look like; that is, the intuitiveness of the above results seems to suggest that the correct theory of backtrackers *should* account for the distinction between causes and background conditions - and this is a very important result with respect to providing core data for debates about backtrackers where there is currently very little.[142]

---

[142] A. Arregui, 'When Modals Meet: Backtracking Counterfactuals,' [unpublished], is the only article I know of that concentrates on anything like a hint towards analysing backtracking counterfactuals.

247

**References, and Appendix:**

- Aristotle, *Physics,* in ed. David Bostock (1999), OUP.

- Armstrong, (1997): *A World of States of Affairs*, Cambridge University Press.

- Bennett, J. (1987): "Event Causation: the Counterfactual Analysis", *Philosophical Perspectives*, 1: p. 367–86.

- Bennett, J. (1988): *Events and their Names*. Indianapolis: Hackett Publishers.

- Campbell, (1990): *Abstract Particulars,* Oxford: OUP.

- Choi, (2005): "Understanding the Influence Theory of Causation: A Critique of Strevens", in *Erkenntins*, July, p. 101-118.

- Coady, (2004): "Pre-empting Pre-emption", in Collins, Hall and Paul eds.

- Collins, J., Hall, E., and Paul, L. (2004): *Causation and Counterfactuals*. Cambridge, Mass: MIT Press.

- Collins, J. (2000): "Pre-emptive Prevention", *Journal of Philosophy*, 97: p. 223–34.

- Davidson, D. (1980), p. 163-80, in *Essays on Actions and Events.*

- ⸺ (1993): 'Thinking Causes', in Heil and Mele, *Mental Causation.*

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

- Dowe, P. & Noordhof, P. eds. (2007), *Cause and Chance, Causation in an indeterministic world.*

- Dretske, (1977): p. 90-7, 'Referring to Events', in French & Uehling & Wittstein (eds.), *Midwest Studies in Philosophy II.*

- Ehring, D (1987):"Causal Relata", *Synthese*, 73: p. 319–328.

- Elga, A. (2000): "Statistical Mechanics and the Asymmetry of Counterfactual Dependence", *Philosophy of Science*, 68 (Supplement): p. 313–324.

- Fine, K. (1975): "Review of Lewis (1973a)", *Mind*, 84: p. 451–8.

- Garfinkel, A. (1981): *Forms of Explanation*. New Haven: Yale University Press.

- Gettier, (1963): 'Is Justified True Belief Knowledge?', *Analysis* 23, p. 121-123.

- Goodman, N.(1947): "The Problem of Counterfactual Conditionals," *Journal of Philosophy*, Vol. 44, No. 5, p. 113-128.

- Hall and Paul, (2006): "Causation and Pre-emption", unpublished.

- Hall, N. (2000): "Causation and the Price of Transitivity", *Journal of Philosophy*, 97: p. 198–222.

- —— (2002): "Non-locality on the Cheap?", *Noûs*,36, p.276–294.

- —— (2004): "Two Concepts of Causation", in Collins, Hall, and Paul (2004), p. 225–76.

- —— (2007): "Structural Equations and Causation", *Philosophical Studies*, 132(1): p. 109–136.

- Hart, H. L. and Honore, A. (1965): *Causation in the Law*. Oxford: Clarendon Press. Second edition (1985).

- Hausman, D. (1998): *Causal Asymmetries*. Cambridge: Cambridge University Press.

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

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

- —— (2007): "Prevention, Pre-emption, and the Principle of Sufficient Reason", *Philosophical Review*, 116(4): p. 495–532.

- Horwich, P. (1987): *Asymmetries in Time*. Cambridge, Mass: MIT Press.

- Hume, D. (1748): *An Enquiry concerning Human Understanding*.

- Kim, J. (1973): "Causes and Counterfactuals", *Journal of Philosophy*, 70: p. 570–72.

- —— (1976): 'Events as Property Exemplification', p.159-77 of Brand and Walton, *Action Theory*.

- Kvart, I. (1986): *A Theory of Counterfactuals*, Ridgeview publishing.

- —— (2001): "Counterexamples to Lewis' 'Causation as Influence'", *Australasian Journal of Philosophy*, 79: 411–23.

- —— (2001): "Lewis's Causation as Influence", in Volume 79, Number 3, September 01, *Journal of Philosophy* p. 409-421 (13).

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

- —— (1973b): "Causation", *Journal of Philosophy*, 70: p. 556–67. Reprinted in his (1986a). Page references to reprinted version.

- —— (1979): "Counterfactual Dependence and Time's Arrow", *Nous*, 13: p. 455–76. Reprinted in his (1986a).

- —— (1980): "A Subjectivist's Guide to Objective Chance", in R. Jeffrey, ed., *Studies in Inductive Logic and Probability: Volume II*, Reprinted in his (1986a).

- —— (1986a): *Philosophical Papers: Volume II*. Oxford: Oxford University Press.

- —— (1986b): "Events", in his (1986a).

- —— (1986c): "Postscripts to 'Causation'", in his (1986a).

- —— (1986d): "Causal Explanation", in his (1986a).

- —— (1986e): *The Plurality of Worlds*. Oxford: Blackwell.

- —— (1999): Papers in Epistemology and Metaphysics. Cambridge Press.

- —— (2000): "Causation as Influence", *Journal of Philosophy*, 97, p.182–97. Reprinted in his (2004a).

- —— (2004a): "Causation as Influence", in Collins, Hall, and Paul (2004), p. 75–106.

- —— (2004b): "Void and Object", in Collins, Hall, and Paul (2004), p. 277–90.

- Lipton, P. (1991): *Inference to the Best Explanation*. London: Routledge.

- Lyon, A. (1967): "Causality", *British Journal for the Philosophy of Science*, 18: 1–20.

- Mackie (1974): *The Cement of the Universe*. Oxford: Oxford University Press. Second edition 1980.

- Maslan, "Degrees of Influence and the Problem of Pre-emption", in *Australasian Journal of Philosophy*, Volume 82, Number 4, December 01, (2004b), p. 577-594 (18).

- Maslen, C. (2004): "Causes, Contrasts, and the Nontransitivity of Causation", in Collins, Hall, and Paul (2004a), p. 341–358.

- Maudlin, T. (2004): "Causation, Counterfactuals, and the Third Factor", in Collins, Hall, and Paul (2004), p. 419–443.

- McDermott, M. (1995): "Redundant Causation", *British Journal for the Philosophy of Science*, 40: p. 523–544.

- Mellor, D. H. (1995): *The Facts of Causation*. London: Routledge.

- —— (2004): "For Facts as Causes and Effects", in Collins, Hall, and Paul (2004), p. 309–324.

- Menzies, P. (1989), p. 59- 83, "A Unified Account of Causal Relata", *Australasian Journal of Philosophy*, 67.

- —— (1999): "Intrinsic versus Extrinsic Conceptions of Causation", in H. Sankey, ed., *Causation and Laws of Nature*, Kluwer Academic Publishers, p. 313–29.

- —— (2004a): "Difference-Making in Context", in Collins, Hall, and Paul (2004), p. 139–80.

- —— (2007): "Causation in Context", in Price and Corry (2007), p. 191–223.

- Mill, J. S. (1843): *A System of Logic*.

- Noordhof, P. (1999):"Probabilistic Causation, Pre-emption, and Counterfactuals", *Mind*, 108: p. 95–125.

- —— (2004): "Prospects for a Counterfactual Theory", in Dowe and Noordhof (2004), p. 188–201.

- Papineau, (1985): 'Causal Asymmetry', *The British Journal for the Philosophy of Science*, 36(3), p. 273-289.

- Paul, L. (2004): "Aspect Causation", in Collins, Hall, and Paul (2004), p. 205–224.

- Pearl, J. (2000): *Causality*. Cambridge: Cambridge University Press.

- Price, H. (1996): *Time's Arrow and Archimedes' Point*. Oxford: Oxford University Press.

- Quine, p. 162-71, 'Events and Reification', in LePore and McLaughlin eds., 1985, *Actions and Events: Essays on the Philosophy of Donald Davidson.*

- Ramachandran, M. (1997): "A Counterfactual Analysis of Causation", *Mind*, 106: p. 263–77.

- Schaffer, J. (2000): "Trumping Pre-emption", *Journal of Philosophy*, 9: p. 165–81.

- —— (2005): "Contrastive Causation", *Philosophical Review*, 114: p. 297–328.

- —— (2007): "The Metaphysics of Causation", *The Stanford Encylopedia of Philosophy*.

- —— (2008): "Causation and Laws of Nature: Reductionism", forthcoming in *Contemporary debates in Metaphysics*.

- Slater, (1992): *The collected papers of Bertrand Russell* vol. 6, p.193-210, "On the notion of cause" by B. Russell.

- Strevens, (2003): "Against Lewis's New Theory of Causation: A Story with Three Morals", in *Pacific Philosophical Quarterly*, Volume 84, Number 4, December, p. 398-412 (15).

- Tooley, (1987): *Causation: A Realist's Approach*, Oxford: Clarendon Press.

- Woodward, J. (1984): "A Theory of Singular Causal Explanation", *Erkenntnis*, 21: p. 31–62.

- Woodward, J. and Hitchcock, C. (2003): "Explanatory Generalizations. Part I: A Counterfactual Account", *Nous*, 37: p. 1–24.

**6. Appendix**

**The Isostructural Analysis of Causation; an analysis with too many epicycles.**

In the following postscripted paper I attempted to analyse causation in terms of some pretty terrifying epicycles. The analysis was demonstrated to be a lot more successful than the prominent analyses of causation that we considered earlier (and its relative successes are plotted in the table I give in the conclusion). I concluded in the below paper that its success would have to stand aside to an analysis that was simpler and more intuitive, and given the analysis I offer in chapter 4 is superior in this regard, the isostructural analysis must stand aside. Another attractive feature of the analysis of chapter 4, in contrast to the isostructural analysis, is that it deals with cases of trumping pre-emption in a preferable way (i.e., in the way Schaffer would have intended, not as a case of symmetrical over-determination as it turns out the isostructural analysis analyses it as). I present the theory below.

I have decided not to put the theory in the main body of the thesis because of its over-complexity. I include the theory in this postscript because, although I now reject it, it stands to show that strategies of 'getting rid of' and 'holding fixed' really are very profitable strategies to think about when considering counterfactual analyses of causation, with the theory furthermore confirming the utility of the strategies I employ in chapter 4. The theory also supports the view that there are many different ways to employ

the latter strategies, with slight differences in details making all the difference in extensional adequacy.

## 1. The Proposed Analysis

In this paper I offer an analysis of causation based on counterfactuals; I term the analysis 'the isostructural analysis of causation', which runs as follows:

C causes E iff

- There is an isostructural duplicate ☐ of a set of actual events ☐ in which Ed. counterfactually depends on Cd.[143]
- C and E are distinct actual events in ☐.

The first condition analyses the causal relation, which I explain as follows:

- E *counterfactually depends* on C iff it's the case that if C hadn't occurred then E wouldn't have occurred.[144]
- A set of events ☐ is an isostructural duplicate of a set of events ☐ iff

---

[143] I use the sufffix 'd.' as an abbreviation of 'duplicate'.

[144] Following Lewis, it's the case that if C hadn't occurred, then E wouldn't have occurred iff the most similar possible worlds in which C does not occur are not E worlds. See Lewis, 'Causation', [1973], for an application of this analysis of counterfactuals to causation.

i) ⬜ and ⬜ are isonomic duplicates.[145]

ii) In ⬜, all chains of dependence between the C and E duplicates only occur between events in ⬜.

iii) What would have changed had the C duplicate not occurred in ⬜, corresponds to the changes made to events other than E had C not occurred, given the set of positive events that stand as pure dependence preventers of E on C had been held fixed.

**Preliminaries:**

**Counterfactual dependence**

Following Lewis, counterfactual conditionals are analysed in terms of a possible worlds semantics. A possible worlds semantics gives truth conditions for counterfactuals in terms of relations between possible worlds.[146] The relation utilised for causal counterfactuals is one of *comparative similarity* between worlds: W1 is said to be *closer* to W0 than W2 if W1 has more relevant properties in common with W0 than W2 does. The relevant properties here are nomological and local, such that the notion of comparative similarity can be given truth conditions as follows:[147]

---

[145] Lewis introduces the terminology of 'isonomic duplicates' into the causal vocabulary in his 'Postscripts to Causation', [1986].

[146] I leave aside the issue of what these possible worlds actually are.

[147] Lewis, [1986d]

W1 is closer to W0 than W2 iff

- As a matter of 1st priority, W1 avoids *widespread* violations of W0's nomological laws more than W2.

- As a matter of 2nd priority, W1 *maximises* the spatio-temporal perfect match of particular matters of fact from W0 more than W2.

- As a matter of 3rd priority, W1 avoids *local* violations of W0's nomological laws more than W2.

- As a matter of 4th priority, W1 secures *approximate* similarity of W0's particular matters of fact more than W2.

The similarity relation also follows three formal constraints:[148]

- It produces a weak ordering, i.e. a connected and transitive relation of worlds in which any worlds can be ordered with respect to their closeness to the actual world.

- The actual world is closest to actuality, resembling itself more than any other world resembles it.

- It is context dependent.[149]

---

[148] Lewis, [1973]

[149] I mention this constraint because it is Lewis' own. The context dependency of the similarity relation can be demonstrated by the following example: Suppose Caesar was in command in the Korean war. Now, we can either hold his military knowledge fixed or his ancient weaponry, such that in one context it is closer to

257

The similarity/closeness relation is then used as a relation between possible worlds to give the truth conditions for counterfactual dependence as follows:

- E counterfactually depends on C iff it's the case that if C had not occurred E would not have occurred iff the closest possible ~C worlds are ~E worlds.

**Isostructural duplicates:**

Isostructural duplicates are given three necessary and together sufficient conditions (listed above), which I explain as follows:

**Condition i)**

The first condition of the analysis for isostructural duplication entails that if a set of events ▯ is an isostructural duplicate of an actual set of events ▯, then they are isonomic duplicates. The conditions for duplication, isonomic duplication, laws of nature, properties, and intrinsic properties, are given as follows:

the actual world to suppose that Julius Caesar would have used the atom bomb if he was in command, and in another context that he would have used catapults. I shall not investigate context independent accounts.

- Two sets of positive events are duplicates iff they have the same intrinsic properties.

- Two sets of positive events are isonomic duplicates iff they have the same intrinsic properties and laws of nature. [150]

- Following the systems approach[151]; laws of nature are theorems of the axiomatisation of the facts that best balance explanatory simplicity and strength. They are also true, non-empty, contingent, fact-supervenient propositions of the form 'All Fs are Gs'.

Given these conditions, it is true that two sets of duplicated positive events are isonomic iff they both occur in nomologically possible worlds. Two sets of duplicated positive events can be isonomic duplicates in one of three ways. Firstly, the two sets of events may be the same indistinct set of events. Secondly, the two sets of events may be distinct sets, and exist in the same world. Thirdly, the two sets of events may be distinct sets, and exist in a distinct but nomologically possible world. Thus, there are three ways in which an isonomic duplicate can feature in my analysis.

**Condition ii)**

The second condition of the analysis for isostructural duplication entails that if a set of events ⬚ is an isostructural duplicate of an actual set of events ⬚, then it's the case that in ⬚,

---

[150] Lewis, [1986], p. 206

[151] Of Mill, Ramsey, and Lewis.

all chains of (counterfactual) dependence between the C and E duplicates only occur between events in □. In short, this condition ensures that there is no event in a chain of dependence between the C and E duplicates that does not feature in the set of events the isonomic duplicate is a duplicate of. The conditions for a chain of dependence are given as follows:

- C & E are linked in a chain of dependence (in which C is the ancestor) iff E depends on D, D depends on D1…. etc., such that Dn depends on C.

C and E may be linked by distinct chains of dependence. For instance, E may depend on F, and F may depend on C, and it may also be the case that E depends on G, and G depends on C (where it may or may not be the case that F depends on G or the converse or not). For □, all distinct chains of dependence between the C and E duplicates must only occur in □.

**Condition iii)**

The third condition of the analysis for isostructural duplication entails that if a set of events □ is an isostructural duplicate of an actual set of events □, then it's the case that what would have changed had the C duplicate not occurred in □, corresponds to the changes made to events other than E had C not occurred, given the set of positive events that stand as pure dependence preventers of E on C had been held fixed. The truth

conditions for the distinction between impure and pure dependence preventers are given as follows:

- An event P impurely prevents dependence of E on C iff P prevents an actual event Q from occurring, such that if P had not occurred, E would be counterfactually dependent on C.

- An event P *purely prevents dependence* of E on C iff some actual events S would form a chain of dependence between C & E (in which C is the ancestor), had P not occurred and S been held fixed.[152]

- If it's the case that had P not occurred and S had been held *fixed*, then S still would occur despite P not occurring (regardless of whether S and P are counterfactually related or not).[153]

**Analysis of Early, Late, and Middle Pre-emption.**

In this section we recall our basic examples of coarse-grained early, late, and middle pre-emption, given in the main body of the thesis.

---

[152] Collins introduces the terminology of 'pure dependence preventers' into the causal vocabulary in his 'Pre-emptive Prevention', [2000].

[153] Yablo introduces the terminology of 'holding fixed' into the causal vocabulary in his 'Advertisement for a Sketch of an Outline of a Proto-theory of Causation', [2004].

Note that in all three cases, E (the shattering) does not counterfactually depend on C (Suzy's throw). The cases of early and middle pre-emption can be analysed by my analysis in two contrasting ways, and late pre-emption by the second way:

*First way*: For the first two cases, take the *actual* world in which the events of Suzy's throw, the flight of the ball, and the shattering occur (such that the isonomic duplicate we take account of in the analysis is the actual world itself). Make sure that what would have changed had the duplicate of Suzy's throw (C) not occurred, corresponds to the changes to the events other than the shattering (E) had her throw (C) not occurred in the actual case, given the trajectory of Billy's ball (the event that stands as the actual pure dependence preventer of E on C) had been held fixed. Given the trajectory of Billy's ball is now held fixed in the actual case, it is true that if C had not occurred, E wouldn't have. Thus the analysis tells us that Suzy's throw (C) is a cause of the shattering (E); the correct analysis.

*Second Way*: For all three cases, take a *different* isonomic duplicate (in this world, or another) of a set of actual events including Suzy's throw, the flight of her ball, and the shattering (such that Billy does not feature in this duplicate). Make sure that what would have changed had the duplicate of Suzy's throw (C) not occurred, corresponds to the changes made to the events other than the shattering (E) had her throw (C) not occurred in the actual case, given the trajectory of Billy's ball (the event that stands as the actual pure dependence preventer of E on C) had been held fixed. The total set of changes to events other than the shattering (E) events are that Suzy's ball would not have flown

towards the window had she not thrown; it would have remained in her hand, and given these changes (to not E events) must correspond to what would have occurred had the duplicate of Suzy's throw (C) not occurred in an isonomic duplicate (which excludes a duplicate of Billy), E (the shattering) would not have occurred. Thus the analysis again tells us that Suzy's throw (C) is a cause of the shattering (E); the correct analysis.

Note five things:

1: Note these two ways of analysing C as a cause of E are different. In the first type of analysis, we can take the actual case (which *includes* Billy and his ball), hold fixed the pure dependence preventers, and we see that E would not occur had C not. In the second type of analysis, we can take a different case which is an isonomic duplicate of a set of events from the actual case (which includes Suzy's throw, the ball's flight, and the shattering, but *excludes* Billy and his ball), make sure that what would have occurred had the duplicate of C not occurred corresponds to what would have occurred to the not E events had C not occurred given the event that stands as the pure dependence preventer had been fixed, and the duplicate of E would not occur had the duplicate of C not occurred. Both types of ways are subsumed by the analysis.

2: Note that the second type of way is the only way we can analyse cases of late pre-emption, given we want an analysis to be able to analyse causation between coarse-grained events. This is because in the first type of analysis which takes the actual case as the isonomic duplicate in question, even if we hold fixed the pure dependence preventers

(the trajectory of Billy's ball in flight towards the window), a coarse-grained version of the shattering still would have occurred; Billy's ball still would have hit the window slightly later than Suzy's would have. Note that this first type of analysis would succeed if we take the causal relata to be fine-grained events, but the first way analyses such cases fine, so we need not.

3. Note that it is necessary to fix the events that stand as the pure dependence preventers in the actual case. Had I not fixed them, then the events that stand as pure dependence preventers in the actual case would have changed had C not occurred (given Billy's ball is minutely gravitationally attracted to Suzy's ball). And given these events would have changed, these changes would have to feature in the world in which the duplicate of C had not occurred given the rest of my analysis; and so E would not counterfactually depend on C, such that C would *not* be analysed as a cause of E; the wrong analysis.

4. Note also that counterfactual dependence is sufficient for dependence in a complete isonomic duplicate, for where an event E counterfactually depends on C, it would also counterfactually depend on C had the set of pure dependence preventers (of which there are none) had been fixed, in the sort of isonomic duplicate required in the analysis.

5. Note that the analysis does not tell us that causation is transitive. This is a virtue for those of us who do not think that causation is transitive; if you think causation is transitive, then I'm happy for you to work that in. As it stands, I think analyses of causation which require transitivity (even to get some of the simplest cases of pre-

emption right), analyse an altogether different causal concept; what it is to be linked to E in a *causal history*,[154] as opposed to what it is to *cause* E simpliciter.

I'll now move on to how the analysis deals with cases of double and trumping pre-emption. In each case, I'll introduce the two different types of examples.

**Analysis of Double Pre-emption**

In this section we recall the case of strengthened over-determination given in the main body of the thesis. For this case we take a different isonomic duplicate (in this world, or another) of Suzy's pressing the light switch (C) such that Billy's pressing (A) does not occur in that isonomic duplicate. Make sure that what would have changed had the duplicate of Suzy's pressing (C) not occurred, corresponds to the changes made to the events other than the light turning on (i.e. to events other than E) had her pressing (C) not occurred in the actual case, given Billy's pressing (A) (the event that stands as the pure dependence preventer of E on C) had been held fixed. The total set of changes has nothing to do with the other over-determiner, and given these changes (to events other than E) must correspond to what would have occurred had the duplicate of Suzy's pressing (C) not occurred in an isonomic duplicate, the light turning on (E) would not have occurred. Thus the analysis tells us that Suzy's pressing (C) is a cause of the light turning on (E) in all cases; the correct analysis.

---

[154] See Lewis, 'Causal Explanation', for some discussion about causal histories.

Have I analysed cases of over-determination correctly? One might think that the light turning on still occurs in the isonomic duplicate - this is not so; what happens had the duplicate of C not occurred, need only correspond to what changes there would be to events *other than E had* C not have occurred in the actual case. So:

- If the fine-grained version of E is over-determined, then what would have occurred had C not occurred in the actual case is that E *would not* have changed, and so E need not feature in the world in which the duplicate of C had not occurred, because only those *changes* to events other than E must feature.

- If the coarse (and not fine) grained version of E is over-determined, then what would have occurred had C not in the actual case is that E *would have* changed, but this changed E need not feature in the world in which the duplicate of C had not occurred, because only those changes to events *other* than E must feature.

**Analysis of Pre-emptive Prevention**

In this section we recall the basic and strengthened examples of pre-emptive prevention outlined in the main body of the thesis.

*Analysis of basic pre-emptive prevention*: Take an isonomic duplicate (that is the actual case, or is a distinct copy) of a set of actual events including the throw of the ball, Suzy's catch, and the presence of the window (it doesn't matter whether Billy features in the

duplicate or not). Make sure that what would have changed had the duplicate of Suzy's catch (C) not occurred, corresponds to the changes made to the not E events had her catch (C) not occurred in the actual case, had the event of Billy standing with his hands in his pockets (the pure dependence preventer of E on C) been held fixed. The total set of changes are that the ball would have flown towards the window had she not caught it, and given these changes (to events other than E) must correspond to what would have occurred had the duplicate of Suzy's catch (C) not occurred in an isonomic duplicate (which might exclude or include a duplicate of Billy), the ball not hitting the window (E) would not have occurred. Thus the analysis tells us that Suzy's catch (C) is a cause of the ball not hitting the window (E); the correct analysis.

*Analysis of the strengthened pre-emptive prevention:* Take an isonomic duplicate (that is the actual case, or is a distinct copy) of a set of actual events including the throw of the ball, Suzy's catch, and the presence of the window (it doesn't matter whether the wall features in the duplicate or not). Make sure that what would have changed had the duplicate of Suzy's catch (C) not occurred, corresponds to the changes made to the events other than E had her catch (C) not occurred in the actual case, given the events of the wall's presence (the pure dependence preventer of E on C) had been held fixed. The total set of changes are that the ball would have bounced off a strong structure had she not caught it, and given these changes (to not E events) must correspond to what would have occurred had the duplicate of C (the catch) not occurred in an isonomic duplicate (which excludes or includes a duplicate of the wall), the ball not hitting the window (E) would

not have occurred. Thus the analysis tells us that Suzy's catch (C) is *not* a cause of the ball not hitting the window (E); the correct analysis.

**Some concluding remarks**

As far as getting the pre-emption cases right, the iso-structural analysis seems to do quite well. That said, the analysis I gave in chapter 4 is a lot easier to understand, and it more effectively takes account of the morals and lessons I suggested at the end of chapter 3. It also deals with cases of trumping in the way that Schaffer originally intended (given that the above analysis analyses cases of trumping as cases of symmetrical over-determination), and so I now reject the above analysis.

**Words total: 61,576**