Does There Exist a Reductive Analysis of Causation, and If so, What is It?

James Read

August 31, 2011

Given that statements of the form 'x caused y' feature extremely frequently in both everyday life and academic disciplines, one might suppose that there exists an uncontroversial reductive analysis of just what is *means* for such an x to cause such a y[1]. However, one does not need progress far in attempting to provide such an analysis to realise that the task is fraught with difficulty, to the extent that some have questioned whether it is possible at all. To reach our own conclusions on this matter though, we should evaluate the various analyses of causation and the difficulties they face for ourselves.

Before we begin this task, some key points in the debate over the nature of causation must be clarified. Firstly, we must distinguish between *type causation* and *token causation*[2]. While the former focuses on general causal claims of the form 'events of type T_1 cause events of type T_2', token causation pertains to specific claims of the form 'event c caused event e.' Since it seems reasonable to believe that an understanding of general causal patterns hinges on an understanding of specific token cases, the contemporary philosophical debate on causation has largely restricted itself to token causation, and we follow this trend.

Secondly, there exists the outstanding issue of the ontological nature of the relata of causal statements (e.g. of the x and y in the statement 'x caused y'): should these be taken to be *events*, or *states of affairs*, or some other form of entity? To answer this question, consider some paradigm cases of causal statements: "The explosion caused the collapse of the building;" "The impact of the cue ball caused the black ball to go into the pocket;" "The chemical caused the plant to

[1] By a reductive analysis, we mean an analysis of the truth-conditions for such a statement not itself employing any causal concepts.

[2] Also called *singular causation*.

wilt." While on the face of it these examples point to a diverse ontology for causal relata (impacts and chemicals, for example, clearly belong to different ontological categories), it is, in fact, natural to specify that all causal relata are *events*. To see this, note that we can easily paraphrase the above three statements into statements concerned with events: "the *event* of the explosion cased the *event* of the building's collapse;" "the *event* of the cue ball impacting with the black ball caused the *event* of the black ball entering the pocket;" "the *event* of the application of the chemical to the plant caused the *event* of the plant's wilting." While there are certainly arguments that can be put forth in favour of taking causal relata to belong to ontological categories other than events - some of which we shall look into in due course - it at least seems a reasonable starting point to take all cases of token causation to be cases of event causation.

Thirdly, modern physics provides strong reasons to believe that the universe is *not* deterministic in nature. Nevertheless, we still suppose the concept of causation has some meaning in our universe. Therefore, any analysis of causation which we provide should be capable of accounting for the fact that the existence of a cause c need not *guarantee* the existence of its effect e: that is, any analysis we provide should be capable of accounting for so-called *chancy causation*[3].

With these clarifications made, we can proceed to consider the diverse array of potential reductive analyses of token causation, and whether any of these are successful. To begin, we consider one basic analysis proposed by David Hume, which we shall call the Humean Analysis of Causation (**HAC**)[4].

HAC: Event c is a cause of event e if and only if (a) c precedes e and (b) c and e are, respectively, events of type T_1 and T_2 such that every event of type T_1 is followed by an event of type T_2 .

On this account, what makes one event a cause of another is that events of the first kind are universally followed by events of the second kind (Hume spoke of the 'constant conjunction' of events of the first kind with events of the second kind[5]). To illustrate, return to our example considered earlier: "the event of the application of the chemical to the plant caused the event of the plant's wilting."

[3] As an example of chancy causation, consider the statement "the presence of the radium cased the Geiger counter to click." This statement may well be true, but the presence of the radium did not *guarantee* the clicking of the Geiger, since it is quite compatible with the probabilistic laws of atomic physics that the radium be present but the counter *not* click.

[4] The specific formulation of this analysis given here is due to Lowe - see Lowe, 'A Survey of Metaphysics.'

[5] See Hume - 'An Enquiry Concerning Human Understanding.'

This account would state that this statement is true because events of the first kind - the application of such chemicals to such plants - are universally followed by events of the second kind - the wilting of such plants.

However, there exist many problems with this account, which ultimately render it untenable. Firstly, there is a problem in determining how the types of events mentioned are to be individuated[6]. Clearly, not every feature of the particular events *c* and *e* can be deemed relevant to determining the types to which they may be said to belong, for if we deem every such feature to be relevant, then we must allow that there are types to which *c* and *e* can be said to belong of which *c* and *e*, respectively, are the sole members. To illustrate in the case of our example, if we take into account every feature of the plant and chemical etc. when individuating the cause and effect events, then these will be so specific that they will be the sole members of event-types T_1 and T_2 respectively. But if this is so, then the truth of clause (b) of **HAC** will follow trivially from the truth of clause (a), implying that any two events where one precedes the other are related as cause to effect, which is absurd. Since there seems to be no clear-cut way to identify types of events, this objection poses a serious problem for **HAC**.

Additionally, **HAC** is unable to overcome the so-called *Problem of Epiphenomena*, which can be demonstrated in this context as follows: suppose that all events of type T_1 are followed by events of type T_2 , and slightly later by events of type T_3. Also suppose that events of type T_2 and T_3 can only occur if preceded by events of type T_1 . Then, given this, it also seems that every event of type T_2 is followed by an event of type T_3. Given this, clauses (a) and (b) of **HAC** are satisfied, and so the analysis erroneously judges events of type T_2 to be causes of events of type T_3, where in fact they are both merely epiphenomena caused by events of type T_1. For example, let events of type T_1 be the type of event of any significant geological disturbance in the Earth's crust (e.g. an earthquake caused by the movement of tectonic plates, or by a nuclear blast). Such disturbances will be followed by P-waves travelling through the Earth (events of type T_2), and shortly after, S-waves travelling through the Earth (events of type T_3). P-waves are always followed by S-waves, but the existence of P-waves travelling through the Earth is not a cause of S-waves travelling through the Earth: in fact they both have a common cause, that being events of type T_1: significant geological disturbances.

HAC is also unable to confront the *Problem of Pre-Emption*, in which in addition to event *c*, another event *d* occurs, such that although *d* is not actually a cause of *e*, if *c* had not occurred, then *d* would still have occurred and would then have

[6] Lowe - 'A Survey of Metaphysics.'

caused *e*. For example, suppose that two assassins line up to kill a victim. The first assassin's shot, *c*, causes the victim's death, *e*, but the second assassin also fires a split-second later than the first and his shot, *d*, although not actually a cause of *e*, would have caused *e* if the first shot, *c*, had not occurred. Here, though it might be true that events of the type to which *d* belongs are universally followed by events of the type to which *e* belongs, it is not the case that *d* qualifies as a cause of *e*, because in fact *e* was caused by *c*, with *d* being irrelevant so long as *e* remained present. Again, **HAC** is unable to accommodate such cases, and so fails.

In light of these problems, we seek a different analysis of causation, turning to that proposed by Mackie[7]. Beginning from the clearly fallacious analysis that 'event *c* is a cause of event *e* if and only if *c* is a necessary and sufficient condition for *e*'s occurring' (this account is evidently wrong: a short circuit may cause a house fire, but it does not *necessarily* cause the house fire, since e.g. the short circuit might have occurred somewhere else; moreover, it is not *sufficient* to cause the house fire, since the presence of oxygen, flammable material, etc. is also required), Mackie refines this to provide the following analysis of causation, which we call Mackie's Analysis of Causation (**MAC**):

MAC: Event *c* is a cause of event *e* if and only if *c* is an *insufficient* but *necessary* part of a condition which is itself *unnecessary* but *sufficient* for *e*.

This analysis is best understood by way of example. In the house fire case discussed previously, we see that the short circuit is a condition that occurred, and that the other conditions which led to the fire, when conjoined with it, form a sufficient condition of which the short circuit is a necessary part. Furthermore, no other sufficient condition of the house's catching fire is also present on this occasion (but it could have been in place of the actual sufficient condition: hence the sufficient condition is unnecessary). Since the short circuit satisfies these criteria, **MAC** tells us that it should be considered a cause of the house fire.

While Mackie admits that refinements to **MAC** may be required, he does assert that "this is an important part of the concept of causation." Nevertheless, **MAC** faces significant problems. Firstly, it seems that the analysis is unable to address the Problem of Epiphenomena discussed previously. To see this, consider again the example concerning geological disturbances, and suppose that an earthquake causes both P-waves and S-waves. Here, it seems that P-waves constitute an insufficient but necessary part of a condition which is itself unnecessary but sufficient

[7] Mackie, 'Causes and Conditions,' in *American Philosophical Quarterly* 2 (1965).

for the S-waves, because there is no way that S-waves could have occurred without the P-waves occurring (given that both are only caused by geological disturbances, whenever geological disturbances occur), but of course the P-waves are insufficient for the S-waves by themselves. However, when conjoined with the earthquake, it is clear that the P-waves form part of a condition which is unnecessary but sufficient for the S-waves (since the S-waves could also have been caused by a nuclear blast etc.). Hence the P-waves seem to satisfy **MAC**, which therefore tells us that they should be considered a cause of the S-waves. But this is evidently wrong, for the same reasons that were given in our discussion of **HAC**: the P-waves do not cause the S-waves; rather they are epiphenomena deriving from the common cause that is the earthquake.

Secondly, Van Fraassen[8] points out that in some cases of indeterministic causation which we know to occur in the real world, there are no sufficient preceding conditions at all, contravening **MAC**. For example, the presence of the radium is what caused the Geiger counter to click, but atomic physics allows a non-zero probability for the counter not clicking at all under the circumstances. Clearly, the conjunction of the radium and the other relevant background conditions is not sufficient for the Geiger counter to click, yet can cause the Geiger counter to click *without* being so sufficient. This seems to constitute a straightforward counterexample to **MAC**. Indeed, in light of the fact that there seems to be no reasonable way to rectify **MAC** to overcome these difficulties, this particular analysis has been broadly rejected in recent times.

The analysis of causation which has receive the most contemporary attention is the so-called Counterfactual Analysis of Causation, brought to prominence by David Lewis[9]. The simplest form of this analysis - which we call the Simple Counterfactual Analysis (**SCA**) - is given below.

SCA: Event c is a cause of event e if and only if (a) c occurs and e occurs and (b) if c had not occurred, then e would not have occurred[10].

Though **SCA** can seem intuitively appealing in its simplicity, it will not do as it stands. To see this, consider cases where one event is part of another event: for example[11], the event of my arm's going up on a certain occasion includes, as a part, the event of my hand's going up on that occasion. Consequently, it seems

[8] Van Fraassen - 'The Scientific Image.'

[9] Lewis, 'Causation,' in *Journal of Philosophy* 70 (1973).

[10] Clause (a) is required to rule out cases of clause (b) being vacuously true.

[11] This example is from Lowe - see Lowe, 'A Survey of Metaphysics.'

that the following counterfactual conditional is true: 'If my arm's going up had not occurred, then my hand's going up would not have occurred.' However, at the same time, the following counterfactual conditional also seems to be true: 'If my arm's going up had not occurred, then my hand's going up would not have occurred.' But if both of these counterfactuals are true, **SCA** implies that each of these events is a cause of the other, which is wrong. To avoid these issues, it seems right to insist that a cause and any one of its effects must be wholly distinct events[12]. As a result, we might amend **SCA** as follows:

SCA': Event *c* is a cause of event *e* if and only if (a) *c* and *e* are wholly distinct events, (b) *c* occurs and *e* occurs, and (c) if *c* had not occurred, then *e* would not have occurred.

Even making this amendment though, **SCA'** faces two further difficulties: firstly, the analysis appears to have difficulty in distinguishing between cause and effect (this is known as the *Problem of Effects*). For example, suppose that a certain bomb contains a reliable mechanism which enables it to be exploded, but only by pressing a certain button, and that, on a certain occasion, the button is pressed and the bomb duly explodes. Here, it does on the face of it seem true to state 'if the explosion of the bomb had not occurred, then the pressing of the button would not have occurred.' But, according to **SCA'**, this means that the explosion of the bomb caused the pressing of the button, where we only want to say the reverse – that the pressing of the button caused the explosion of the bomb.

Secondly, like **HAC** and **MAC**, **SCA'** has difficulty when faced with the Problem of Epiphenomena. To see this, suppose that we again have a bomb fitted with a reliable mechanism which enables it to be exploded only by pressing a certain button, but that the mechanism also enables a warning light to be activated, once more only by pressing the button. The button is pressed and - either simultaneously or in sequence - the warning light flashes and the bomb explodes. We want to say that the pressing of the button caused both the explosion of the bomb and the flashing of the light, but it seems that **SCA'** commits us to saying, in addition, that the explosion of the bomb caused the flashing of the light and vice versa[13], which is clearly mistaken.

[12] This raises the issue of what we mean by describing two events as being 'wholly distinct'. The correct thing to say is that two events are wholly distinct if and only if there is no event which is a common part of both of them.

[13] Since the counterfactuals 'If the flashing of the light had not occurred, the explosion of the bomb would not have occurred and 'If the explosion of the bomb had not occurred, the flashing of the light had not occurred' appear true.

To resolve these issues, Lewis makes the crucial stipulation that the counterfactual conditionals appearing in **SCA'** should not be so-called *backtracking counterfactuals*. To see what is meant by this, we first make clear that *in evaluating the truth or falsity of the counterfactual conditional appearing in* ***SCA'****, we should consider whether or not, in the possible worlds in which c does not occur but everything else happens just as in the actual world up to the time of c's occurrence, e also occurs*. In the case of our bomb example, the counterfactual conditional 'If the explosion of the bomb had not occurred, then the pressing of the button would not have occurred' turns out to be false according to this method of evaluation, because the explosion of the bomb occurs after the pressing of the button, so if we hold fixed everything else that happens up until the bomb's exploding, the pressing of the button will have occurred anyway. Consequently, once we evaluate counterfactuals in **SCA'** in this way, we can rule out as false these problematic backtracking counterfactuals which take the form 'If x had happened at time t, then y would have been different at a time earlier than t.'

Indeed, once this stipulation is made we can also avoid the second difficulty, because this allows us to evaluate as false the counterfactual conditional 'If the flashing of the light had not occurred, then the explosion of the bomb would not have occurred,' even if we assume that the flashing of the light occurred before the explosion of the bomb. To see why, note that if we consider the possible worlds in which the flashing of the light does not occur but everything else happens just as in the actual world up to the time at which the flashing light occurs, we find that such a world is one in which the causal processes leading to the explosion of the bomb are already under way at the time at which the light fails to flash, so that the explosion still occurs.

So, once it is stipulated that the counterfactuals appearing in **SCA'** must not be backtracking counterfactuals, this analysis can avoid the Problem of Effects and the Problem of Epiphenomena. Nevertheless, **SCA'** is still presented with significant problems when faced with the Problem of Pre-Emption discussed previously, in which another event d occurs, such that although d is not actually a cause of e, if c had not occurred, the d would still have occurred and would then have caused e. The reason for this is obvious: in such cases, clause (c) of **SCA'** is not satisfied, as if c had not occurred, e would still have occurred (as it would have been caused by d). Hence, c causes e in spite of not satisfying **SCA'**.

How might one save **SCA'** from this difficulty? One attempt to do this attempt (originally proposed by Lewis) would be to argue that if c (the actual cause of e) had not occurred, then the event that d would have caused would *not* have been *e*, but rather some numerically distinct, though very similar, event. This approach takes events to be highly *fragile*: if conditions were only slightly different, then a numerically different effect - and not simply the same effect altered in some way - would have occurred. In fact though, while the fragility thesis might initially appear tempting, there are many reasons to reject it.

Firstly, we are usually quite happy to say that an event might have been slightly delayed, or that it might have differed in some of its contingent aspects. Given that d may cause an effect e' only infinitesimally different from the effect e caused by c, it seems that, if the fragility thesis is to proceed, it must maintain that no event could have been different from the way it is, for if it were to so differ, it would have been a different event. That is, every event has all its features *necessarily*. But this contrasts with our ordinary way of speaking, where there is flexibility in the properties to have the same event to possess[14].

Furthermore, on the fragility thesis, all manner of irrelevant things which we would not ordinarily count among the causes of the effect can be expected to make some slight difference to its time and manner. For example, if a neutrino happens to pass through a person's body as he dies, we must count the explosion at the center of the sun which produced the neutrino as a cause of his death, because on the assumption that his death is taken as a very fragile event, he would not have died his actual death with a neutrino passing through his body if the explosion had not occurred. If we heed still smaller differences, almost everything that precedes an event will be counted among its causes. So by adopting the fragility strategy, in whichever form, we open the gate to a flood of spurious causes.

Given these problems, a different resolution to the Problem of Pre-Emption in the context of **SCA'** seems in order. Another way of dealing with this problem is to propose a revised analysis of causation, which we call the 'Complex Counterfactual Analysis' (**CCA**)[15].

CCA: Event c is a cause of event e if and only if (a) c and e are wholly distinct events, (b) c occurs and e occurs, and (c) a chain of counterfactually dependent events links c to e.

What clause (c) of **CCA** means is this: there was a finite sequence of actually occurring events, with c being the first member of the sequence and e the last,

[14] For instance, we suppose that the Battle of Hastings could have occurred a day later than it did, but it would still have been the same event.

[15] This account is again due to Lewis - see Lewis, 'Causation,' in *Journal of Philosophy* 70 (1973).

such that each member of the sequence is counterfactually dependent upon the immediately preceding member of the sequence, where the *counterfactual dependence* of e upon c is defined as follows: if c had not occurred then e would not have occurred. To illustrate **CCA**, suppose that there is a sequence of events of this kind which possesses just three members, with c being the first member and e being the last, and call the intermediate event of the sequence x. Then what is required by clause (c) above is that x should be counterfactually dependent upon c and e should be counterfactually dependent on x. This in turn means that the following two counterfactual conditionals should be true: 'If c had not occurred, then$x$ would not have occurred' and 'If x had not occurred, then e would not have occurred.'

How, then, is **CCA** supposed to overcome the Problem of Pre-Emption? The clearest answer to this comes from Lewis himself: "As far as causal dependence goes, there is no difference: e depends neither on c nor d. If either one had not occurred, the other would have sufficed to cause e. So the difference must be that, thanks to c there is no causal chain from d to e; whereas there is a causal chain of two or more steps from c to e. Assume for simplicity that two steps are enough. Then e depends causally on some intermediate event f, and f in turn depends on c."[16] So clearly, in order for **CCA** to succeed, it must posit at least one intermediate event between cause and effect. For example, in our Pre-Emptive assassin case considered earlier, we may say that the event of the first assassin firing his gun c caused the intermediate event f of his bullet hitting the victim, which in turn cased the death of the victim e. Since there is clearly a chain of causal dependence here, we can say that c caused e. However, we can equally use this account to accommodate our intuition that the event of the second assassin firing his gun d does not cause e: this is so since d causes the intermediate event of the second assassin's bullet hitting the victim g - but since this occurs after f, it is clear that e does not counterfactually depend on g, and so by **CCA**, d does not cause e. Hence **CCA** shows itself to be an intuitive and compelling counterfactual analysis of causation capable of dealing with the standard problem cases.

However, **CCA** is not yet home and dry, for while the analysis is able to overcome standard cases of Pre-Emption, a new difficulty is raised in so-called 'Late Pre-Emption'. These cases are very similar to the Pre-Emptive cases already considered, though here the only effect of event c that prevents the completion of a chain of counterfactually dependent events linking the event d to event e is the event e itself (recall that in the previous Pre-Emption case, it was not e but f that

[16] Lewis, 'Causation,' in *Journal of Philosophy* 70 (1973).

prevented a causal chain from d to e). This is again made clearer by way of example: “Billy and Suzy throw rocks at a bottle. Suzy throws first, or maybe she throws harder. Her rock arrives first. The bottle shatters. When Billy’s rock gets to where the bottle used to be, there is nothing there but flying shards of glass. Without Suzy’s throw, the impact of Billy’s rock on the intact bottle would have been one of the final steps in the causal chain from Billy’s throw to the shattering of the bottle. But, thanks to Suzy’s Pre-Empting throw, that impact never happens.”[17]

The problem now is that, in such a case, it seems that there will not be a complete chain of counterfactually dependent events linking the *actual* cause of e, event c, to event e, so that **CCA** will not license us to describe c as being a cause of e. Why? Because if the penultimate event in the putative chain of counterfactually dependent events linking c to e had not occurred, the occurrence of e would not have prevented the completion of such a chain linking d to e, so that e would still have occurred. Hence, e is not counterfactually dependent upon the penultimate event in question, so that these two events do not in fact belong to a chain of counterfactually dependent events linking c to e. To take the example above, **CCA** cannot explain the judgment that Suzy’s throw was the actual cause of the shattering of the bottle, for there is no causal dependence between Suzy’s throw and the shattering, since even if Suzy had not thrown her rock, the bottle would have shattered due to Billy’s throw. Nor is there a chain of stepwise dependences running cause to effect, because there is no event intermediate between Suzy’s throw and the shattering that links them up into a chain of dependences. Take, for instance, Suzy’s rock in mid-trajectory. Certainly, this event depends on Suzy’s initial throw, but the problem is that the shattering of the bottle does not depend on it, because even without it the bottle would still have shattered because of Billy’s throw.

Late Pre-Emption presents a genuine and serious counterexample to **CCA**. As a consequence, a further generation of counterfactual theories of causation has developed, in an attempt to correctly account for these cases. The two most prominent accounts here are Lewis’ 2000 counterfactual account of causation[18] (which we call ’Lewis’ New Account’ (**LNA**)), and the ’Structural Equations Account’

[17] This is a standard example from Lewis - see Lewis, ’Causation as Influence,’ in *Journal of Philosophy* 97 (2000).

[18] Discussed in, for example, Lewis, ’Causation as Influence,’ in *Journal of Philosophy* 97 (2000).

(**SEA**) advocated by, for example, Hitchcock[19] and Woodward[20].

Beginning with **LNA**, the key idea is that of an *alteration* of an event. This is an actualised or unactualised event that occurs at a slightly different time or in a slightly different manner from the given event. An alteration is, by definition, a very fragile event that could not occur at a different time, or in a different manner without being a different event. Using the idea of an alteration defined in this way, Lewis then defines a notion of *influence* as follows: where c and e are distinct events, c influences e if and only if there is a substantial range of c_1, c_2, $\ldots$ of different not-too-distant alterations of c (including the actual alteration of c) and there is a range of e_1, e_2, $\ldots$ of alterations of e, at least some of which differ, such that if c_1 had occurred, e_1 would have occurred, and if c_2 had occurred, e_2 would have occurred, and so on. Given this, **LNA** can be stated as follows[21]:

LNA: Event c is a cause of event e if and only if there is a chain of stepwise influence from c to e.

As well as being able to handle the Problem of Effects, Problem of Epiphenomena, and Problem of Pre-Emption, **LNA** is also capable of addressing cases of Late Pre-Emption. To see this, reconsider the example involving Billy and Suzy. The theory is supposed to explain why Suzy's throw, and not Billy's throw, is the cause of the shattering of the bottle. If we take an alteration in which Suzy's throw is slightly different (the rock is lighter, or she throws sooner), while holding fixed Billy's throw, we find that the shattering is different too. But if we make similar alterations to Billy's throw while holding Suzy's throw fixed, we find that the shattering is unchanged. Hence the theory correctly judges that Suzy's throw is a cause of the bottle shattering, while Billy's throw is not.

As Menzies points out though[22], there is reason to doubt whether **LNA** handles cases of Late Pre-Emption completely satisfactorily. In the above example, Billy's throw has some degree of influence on the shattering of the bottle. For if Billy had thrown his rock earlier (so that it preceded Suzy's throw) and in a different manner, the bottle would have shattered earlier and in a different manner. In response to these points, Lewis must say that these alterations of the events

[19] Hitchcock, 'The Intransitivity of Causation Revealed in Equations and Graphs,' in *Journal of Philosophy* 98 (2001).

[20] Woodward, 'Making Things Happen: A Theory of Causal Explanation.'

[21] This exact formulation of **LNA** is due to Menzies - see Menzies, 'Counterfactual Theories of Causation,' in *The Stanford Encyclopedia of Philosophy*.

[22] Menzies, 'Counterfactual Theories of Causation,' in *The Stanford Encyclopedia of Philosophy*.

are too distant to be considered relevant[23]. But it is unclear why this ruling out on grounds of relevance should always be warranted: if we suppose that Billy's stone would have hit the bottle only a very short time after Suzy's stone actually did, then there does seem to be a close alteration of Billy's throw in which it hits the bottle before Suzy's throw, and so smashes the bottle. In order for Lewis' response to this objection to go through then, some metric of distance in alterations is required, and the burden is on the advocates of **LNA** to provide this.

Additionally, it has been argued that the new theory generates a great number of spurious instances of causation, since **LNA** implies that any event that influences another event in the manner defined above counts as one of its causes. But commonsense is more discriminating about causes. To take an example, rain in December delays a forest fire; if there had been no December rain, the forest would have caught fire in January rather than when it actually did in February. The rain influences the fire with respect to its timing, location, rapidity, and so forth. But commonsense denies that the rain was a cause of the fire, though it allows that it is a cause of the delay in the fire. Clearly then, **LNA** does still face difficulties, at least when formulated as above.

Let us see whether **SEA** fares any better. This analysis describes the causal structure of a system of events in terms of a causal model of the system, which is identified as an ordered pair $\langle V, E \rangle$, where V is a set of variables and E a set of so-called *structural equations* stating relations among the variables. The variables in V describe the different possible states of the system in question[24]. Again, this approach is best illustrated by way of example: let us formulate a causal model to describe the system exemplified in the example of Late Pre-Emption involving Billy and Suzy. We describe the system using the following set of variables:

- $BT = 1$ if Billy throws a rock, 0 otherwise;
- $ST = 1$ if Suzy throws a rock, 0 otherwise;
- $BH = 1$ if Billy's rock hits the bottle, 0 otherwise;
- $SH = 1$ if Suzy's rock hits the bottle, 0 otherwise;
- $BS = 1$ if the bottle shatters, 0 otherwise.

[23] That is, belong to possible worlds not sufficiently similar to the actual world.

[24] While they can take any number of values, in the simple examples to be considered here the variables are binary variables that take the value 1 if some event occurs and the value 0 if the event does not occur.

The structural equations in a model describe the dynamical evolution of the system being modelled. There is a structural equation for each variable. The form taken by a structural equation for a variable depends on which kind of variable it is: *exogenous variables* take values which are determined by factors outside the model; their structural equations take the form $Y = y$, which simply states the actual value of the variable[25]. By contrast, *endogenous variables* have values which are determined by factors within the model; their structural equations take the form $Y = f(X_1, \ldots, X_n)$, which states how the value of the variable is determined by the values of the other variables. The equation for an endogenous variable encodes a set of counterfactuals of the following form: 'If it were the case that $X_1 = x_1, X_2 = x_2, \ldots$, $X_n = x_n$, then it would be the case that $Y = f(x_1, \ldots, x_n)$'.

As this form of counterfactual suggests, the structural equations are to be read from right to left: the antecedent of the counterfactual states possible values of the variables X_1 to X_n and the consequent states the corresponding value of the endogenous variable Y. There is a counterfactual of this kind for every combination of possible values of the variables X_1 through to X_n. An important feature of the structural equations for endogenous variables is that they must be complete in the sense that the equation for a variable Y must express the value of Y given the values of all and only the variables X_i on which it counterfactually depends, for a given combination of the values of those variables.

Now that we understand how these structural equations work, we can proceed further with our example: consider the set of structural equations that might be used to model the example of Billy and Suzy. Given the variables listed above, the structural equations might be stated as follows[26]:

- $ST = 1$;
- $BT = 1$;
- $SH = ST$;
- $BH = BT \wedge \sim SH$;
- $BS = SH \vee BT$.

[25] The notation is that capital letters represent variables; lower case letters the values taken by variables.

[26] I follow Menzies in employing these particular structural equations - see Menzies, 'Counterfactual Theories of Causation,' in *The Stanford Encyclopedia of Philosophy*.

In these equations logical symbols are used to represent mathematical functions on binary variables: $\sim X \equiv 1 - X$; $X \vee Y \equiv \max\{X,Y\}$; $X \wedge Y \equiv \min\{X,Y\}$. The first two equations above state the actual values of the exogenous variables ST and BT. The third equation encodes two counterfactuals, one for each possible value of ST. It states that if Suzy threw a rock, her rock hit the bottle; and if she didn't throw a rock, her rock didn't hit the bottle. The fourth equation encodes four counterfactuals, one for each possible combination of values for BT and $\sim SH$. It states that if Billy threw a rock and Suzy's rock didn't hit the bottle, Billy's rock hit the bottle; but didn't do so if one or more of these conditions was not met. The fifth equation encodes four counterfactuals, one for each possible combination of values for SH and BH. It states that if one or other (or both) of Suzy's rock or Billy's rock hit the bottle, the bottle shattered; but if neither rock hit the bottle, the bottle didn't shatter.

The structural equations directly encode counterfactuals. However, some counterfactuals that are not directly encoded can be derived from them. Consider, for example, the counterfactual 'If Suzy's rock had not hit the bottle, the bottle would still have shattered.' As a matter of fact, Suzy's rock did hit the bottle. But we can determine what would have happened if it hadn't done so, by replacing the structural equation for the endogenous variable SH with the equation $SH = 0$, keeping all the other equations unchanged. So, instead of having its value determined in the ordinary way by the variable ST, the value of SH is set 'miraculously.' After this operation, the value of the variable BS can be computed and shown to be equal to 1: given that Billy had thrown his rock, his rock would have hit the bottle and shattered it. So this particular counterfactual is true[27]. In general, to evaluate a counterfactual, say 'If it were the case that $X_1, \ldots, X_n$, then ...,' one replaces the original equation for each variable X_i with a new equation stipulating its hypothetical value while keeping the other equations unchanged; one then computes the values for the remaining variables to see whether they make the consequent true.

So far so good, but in what way does this approach yield an analysis of causation? To answer this, we must first recognise that this technique of replacing an equation with a hypothetical value set by a 'surgical intervention' again enables us to capture the notion of counterfactual dependence between variables, though in slightly different terms to the way in which this was defined before: A variable Y *counterfactually depends* on a variable X in a model if and only if it is actually

[27] Significantly, this procedure for evaluating counterfactuals directly reflects Lewis' non-backtracking interpretation of counterfactuals: the surgical intervention that sets the variable SH at its hypothetical value but keeps all other equations unchanged is similar in its effects to Lewis' realising the counterfactual antecedent while preserving the past.

the case that $X = x$ and $Y = y$ and there exist values $x' \neq x$ and $y' \neq y$ such that replacing the equation for X with $X = x'$ yields $Y = y'$[28]. Additionally, we define a *route* between two variables X and Z in the set V to be an ordered sequence of variables $\langle X, Y_1, \ldots, Y_n, Z\rangle$ such each variable in the sequence is in V and is a parent of its successor in the sequence (X is a *parent* of the endogenous variable Y if and only if the variable X features as an argument on the right-hand side of the structural equation for Y). A variable Y is *intermediate* between X and Z if and only if it belongs to some route between X and Z.

The route $\langle X, Y_1, \ldots, Y_n, Z\rangle$ is *active* in the causal model $\langle V, E\rangle$ if and only if Z depends counterfactually on X within the new system of equations E' constructed from E as follows: for all Y in V, if Y is intermediate between X and Z but does not belong to the route $\langle X, Y_1, \ldots, Y_n, Z\rangle$, then replace the equation for Y with a new equation that sets Y equal to its actual value in E. (If there are no intermediate variables that do not belong to this route, then E' is just E.) This definition generalises the informal idea sketched in the example of Suzy and Billy. There is an active causal route going from Suzy's throwing her rock through her rock hitting the bottle to the bottle shattering: when we hold fixed Billy's rock not hitting the bottle, which is the actual value of the only intermediate variable *BH* that is not on this route, we see that the bottle's shattering counterfactually depends on Suzy's throwing her rock. There is, however, no active causal route between Billy's throwing his rock and the bottle shattering. In terms of the notion of an active causal route, we are now in a position to provide an analysis of token causation:

SEA: Event c is a cause of event e if and only if (a) c and e are wholly distinct events, (b) X and Z are binary variables whose values represent the occurrence and non-occurrence of these events, (c) there is an active causal route from X to Z in an appropriate causal model $\langle V, E\rangle$[29].

A crucial notion in this definition is that of an *appropriate* model. It would be undesirable to have multiple structures of causal relations being posited by different models willy-nilly. So we insist that causal relations are revealed only by 'appropriate' models. Hitchcock mentions a number of criteria for appraising whether

[28] The similarity with the previous definition is obvious, since explicitly in terms of counterfactuals this new definition would read: 'If X were x' then Y would be y', for some x', y'.'

[29] As mentioned previously, this has been simplified to only accommodate binary variables. For the generalised account, see Hitchcock - 'The Intransitivity of Causation Revealed in Equations and Graphs,' in *Journal of Philosophy* 98 (2001).

a model is appropriate[30], the most important being that the structural equations posited by the model must not imply any false counterfactual. Additionally, Hitchcock defines a notion of a *weakly active route*, there being a weakly active route between X and Y just when Y counterfactually depends on X under the freezing of some possible, though not necessarily actual, values of the variables that are not on the route from X to Y. In the case of Billy and Suzy, it is evident that there is an active causal route between Suzy's throw and the bottle shattering, and a weakly active causal route between Billy's throw and the bottle shattering, indicating that if conditions had been different (i.e. if some structural variables had taken different values - namely if $SH = 0$), Billy's throw would indeed have caused the bottle to shatter.

Clearly, **SEA** is capable of overcoming Late Pre-Emption cases. Still, problems have recently arisen with the analysis which are only now being addressed. Consider one example, due to Menzies[31]: Suppose an assassin puts poison in the king's coffee. The bodyguard responds by pouring an antidote in the king's coffee. If the bodyguard had not poured the antidote in the coffee, the king would have died. On the other hand, the antidote is fatal when taken by itself, and if the poison had not been poured in first, it would have killed the king. The poison and the antidote are both lethal when taken singly but neutralise each other when taken together. In fact, the king drinks the coffee and survives. The objection is that the clauses of **SEA** are satisfied with respect to the assassin pouring poison in the coffee (c) and the king surviving (e), erroneously suggesting that the assassin putting the poison into the coffee caused the king to survive. To see this, consider the most appropriate causal model $\langle V, E\rangle$for this case:

- $A = 1$if the assassin pours poison into the king's coffee, 0 otherwise;
- $G = 1$if the bodyguard responds by pouring antidote into the coffee, 0 otherwise;
- $S = 1$if the king survives, 0 otherwise.

...Employing the following structural equations:

- $A = 1$;

[30] Hitchcock - 'The Intransitivity of Causation Revealed in Equations and Graphs,' in *Journal of Philosophy* 98 (2001).

[31] Menzies, 'Counterfactual Theories of Causation,' in *The Stanford Encyclopedia of Philosophy*.

- $G = A$;
- $S = (A \wedge G) \vee (\sim A \wedge \sim G)$

Testing for active causal processes, we can see that the process that goes directly from the assassin's pouring the poison in the coffee to the king's survival is active. Holding fixed the fact that the bodyguard poured the lethal antidote into the coffee, we note that the king would not have survived if the assassin had not put the poison in the coffee first. So the theory licenses the verdict that the assassin's pouring in the poison caused the king to survive. However, it seems reasonable to say that this is a mistaken causal verdict: putting poison in the king's coffee is exactly the kind of thing that is likely to *kill* the king, not cause him to *survive*. Since there is no other appropriate model which gets the correct result here, there are evidently problems for **SEA**. While revisions to the analysis have been proposed in an attempt to circumvent this issue, they are not only fledgling but also highly complicated; I therefore omit a detailed discussion.

Up to this point, all the theories of causation considered have been deterministic in nature, speaking only of the occurrence or non-occurrence of an effect based on the presence of a cause, and not of any *probabilistic* dependence between cause and effect. However, as already mentioned, it is desirable to possess an analysis of causation which is capable of operating in a chancy universe. As a result, let us consider theories of causation which would be able to accommodate chance, for it might be that if we can accommodate this from the start, the account of causation which we ultimately obtain is more successful. Indeed, once the requirement for probabilistic dependence is laid down, a new idea for an analysis of causation quickly becomes apparent: we call this the Basic Probabilistic Analysis of Causation (**BPA**).

BPA: Event c is a cause of event e if and only if the occurrence of c raises the probability of the occurrence of e by some amount.

One way to interpret 'the occurrence of c raises the probability of the occurrence of e by some amount' is as follows: The conditional probability of e's occurring given the occurrence of c was higher than the conditional probability of e's occurring given the non-occurrence of c. However, there are problems in taking this route: we cannot intelligibly talk of the conditional probability of e's occurring given the non-occurrence of c, so it seems we shall have to talk instead of the conditional probability of e's occurring given the occurrence of some alternative event distinct from c, and compare this with the conditional probability of e's occurring given the occurrence of *c*. This other event will, evidently, have to be some event whose occurrence was incompatible with the occurrence of *c*, so as to exclude the possibility of both events occurring. However, in any given case we shall be able to envisage many other events whose occurrence was incompatible with the occurrence of event *c*, but which could have occurred instead of *c*. How are we to decide? Different choices may yield different verdicts as to whether *c* was a cause according to **BPA**, so it seems that this approach cannot yield a determinate answer as to whether *c* caused *e*.

As an alternative, one might seek to combine **BPA** with the counterfactual analyses discussed previously[32] - to do this, we first define the notion of *probabilistic dependence*: if *c* and *e* are wholly distinct events, then *e* probabilistically depends on *c* just in case if *c* were to occur, the chance of *e*'s occurring would be *x*; and if *c* were not to occur, the chance of *e*'s occurring would be *y*, where *x* is significantly greater than *y*. Like counterfactual dependence in our discussion before, probabilistic dependence is not the same as causation, for it is possible to have causation without probabilistic dependence in, for example, cases of Pre-Emption[33]. Still, we can employ the notion of probabilistic dependence in order to provide an analysis of causation as follows: first, let us state that a finite sequence of events $\langle a, b, c, \ldots \rangle$ is a *chain of probabilistically dependent events* if and only if *b* probabilistically depends on *a*, *c* probabilistically depends on *b*, and so on. Given this, we can (by analogy with **CCA**) provide a Complex Probabilistic Account of Causation (**CPA**):

CPA: Event *c* is a cause of event *e* if and only if (a) *c* and *e* are wholly distinct events, (b) *c* occurs and *e* occurs, and (c) a chain of probabilistically dependent events links *c* to *e*.

This account allows us to overcome the Pre-Emption cases discussed previously if we posit the intermediate events of the first assassin's bullet hitting the victim and the second assassin's bullet hitting the victim, in a manner directly analogous to our discussion of **CCA**. However, like **CCA**, the analysis is not able to overcome problems of Late Pre-Emption, and indeed the example of Billy and Suzy serves to illustrate this once more. In fact, this nips in the bud analyses which attempt to combine the probabilistic account with counterfactual accounts - there seems to be no easy way to progress from this point. While Pearl has suggested a revision

[32] See, for example, Lewis, 'Postscripts to 'Causation',' in *Philosophical Papers, Volume II.*

[33] Our standard assassin Pre-Emption case serves to illustrate this, for if the first assassin were not to fire his gun, the probability of the victim dying would remain more or less the same, since the second assassin would shoot and hit the victim in this case.

of **SEA** capable of accommodating chancy causation[34], this approach has not yet been considered in detail, and so it is too early to pronounce upon its success.

Additionally, there do unfortunately exist reasons to doubt the success of *any* account of causation based upon the idea of probability-raising[35]. For example, suppose that two gunmen are shooting at a vase. Each one has a fifty percent chance of hitting the vase, and each one shoots independently, so the probability that the vase shatters is 0.75. As it happens, the first gunman's shot hits the vase, but the second gunman misses. In this example, the second gunman shot at the vase, his shooting increased the probability that the vase would shatter (from 0.5 to 0.75), and the vase did in fact shatter. Nonetheless, it seems clear that we should *not* say that the second gunman's shot caused the vase to shatter. So here we have an apparent counterexample to probability-raising theories of causation: the second shot significantly increased the probability that the vase would shatter, but it did not cause the vase to shatter. In an extended discussion, Hitchcock maintains that it is unclear whether counterexamples of this kind are successful; but if they are, they evidently pose a sizable problem for any such analysis of causation.

In light of our discussion up to this point, it is hard to reach any positive conclusions as to whether there exists a successful reductive analysis of causation, for all the analyses which we have considered have faced difficulties. Still, it might at this point be suggested that the reason for the difficulties encountered is due to the fact that causal relata are not, in fact, events, as was previously assumed. Perhaps the most well-known alternative analysis of causation is that due to Salmon[36], which focuses on *extended causal processes*, by which is meant any spatio-temporally continuous series of events (e.g. a particle travelling through some medium).

The class of extended causal processes can be divided into those which are *genuine causal processes*, and those which are merely *pseudo-processes*. To note the difference, consider a car moving along a road. As the car moves, its shadow moves along the road too. The series of events in which the car occupies successive points on that road is a genuine causal process, while the movement of the shadow is merely a pseudo-process, because the position of the shadow at later times is not caused by its position at earlier times. Here, we need only consider

[34] Pearl, 'Structural Equations and Probabilistic Causality.'

[35] Hitchcock, 'Do All and Only Causes Raise the Probability of Effects?', in Collins, Hall, and Paul, *Causation and Counterfactuals*.

[36] Salmon, 'Causality: Production and Propagation,' in *Proceedings of the Biennial Meeting of the Philosophy of Science Association*, vol. 2 (1980)

genuine causal processes[37]. What tie continuous causal processes together are, on Salmon's theory, causal *interactions*. These interactions are the 'forks' that combine all those causal processes into a causal structure. For example, suppose that the cue ball on a billiard table travels towards, collides with, and pockets the black ball. Before the collision of the balls, there exist two relevant causal processes: the cue ball travelling towards the black, and the black remaining stationary. After the collision, we have the black ball travelling towards the pocket, and the cue ball travelling off in some other direction. What connects the two pre-collision causal processes to the two post-collision causal processes is the causal interaction of the collision.

Salmon gives a detailed account of the nature of these causal interactions, though in fact we need not discuss this to already see that his theory faces significant problems. Firstly[38], we often speak of causation in the context of physical theories which involve action at a distance, for example classical electrodynamics or Newtonian mechanics. While it may be assumed that all such theories have, or will, be superseded by theories which do not employ action at a distance (for example, electrodynamic phenomena have since been accounted for via the mechanism of photon exchange), there is no reason to suppose they are so replaceable. Indeed, if this turns out to not be the case, would that mean that these theories could not be deployed in causal explanations, because according to them there is no continuous processes which describes, for example, how the presence of the asteroid can *cause* the perturbation in the planetary orbit? This would be an unacceptable conclusion to draw, since we ordinarily suppose that we can employ these theories unproblematically in causal explanations.

Additionally, there are further reasons that mitigate against requiring that causation always involve connection via continuous processes[39]. Suppose that an assassin fires at a victim, but a secret service agent jumps in front of the assassin's bullet, allowing himself to be hit. The agent's action caused the target to remain alive, even though there are no processes connecting the agent and the target (except for irrelevant ones such as sound waves and photons). It is not easy to see how an account of causation which relies on every case being analysable in terms of continuous causal processes can overcome problems such as this; hence it is

[37] Salmon proposes a method of distinguishing causal processes from pseudo-processes based on their ability to transmit a mark, though we simply assume that genuine causal processes *can* be distinguished from mere pseudo-processes in some way.

[38] This objection is from Van Fraassen - see Van Fraassen, 'The Scientific Image.'

[39] This objection is from Hitchcock - see Hitchcock, 'Do All and Only Causes Raise the Probability of Effects?' in Collins, Hall, and Paul (eds.), *Causation and Counterfactuals*.

hard to see how Salmon's analysis can succeed as it stands.

There do not presently exist any unproblematic alternative accounts of causation taking causal relata to be constituted by some entity other than events. While some, such as Lowe[40] and Carroll[41], have taken this as yet further evidence that causation is an irreducible and fundamental concept, at the present time this seems a hasty and pessimistic conclusion to draw: given the furtive research in this field still in progress, and the technical power of theories such as **SEA** (which seems the least problematic of the analyses discussed here, and that which has the greatest potential for future development), it seems far too early to simply assume that this research will not prove fruitful. Indeed, taking causation to be irreducible can only be a last-ditch effort, and until every possible reductive analysis is thoroughly explored, we cannot reasonably follow this route. Moreover, even supposing irreducible causation, there is great value in seeking reductive analyses which better approximate the true concept, for as was mentioned at the outset, rigorously establishing whether one event was the cause of another is of crucial importance both in everyday life and academic disciplines.

[40] Lowe, 'A Survey of Metaphysics.'

[41] Carroll, 'Nailed to Hume's Cross?' in Sider, Hawthorne, and Zimmerman, *Contemporary Debates in Metaphysics*.