The Emergence of Causation

J. Dmitri Gallow [*,†]

I would like to foreground a division between philosophical theories of singular, token, or actual causation. In one camp are those theories according to which the causal relations of our day-to-day, macroscopic lives, as well as the causal relations investigated by the special sciences, ultimately reduce to causal relations between microphysical events—that the causal relation between Zimbabwe's monetary policy and its hyperinflation is in principle reducible to causal influence between the fundamental particles realizing those events. Following Peter Menzies,[1] call the members of this camp *causal reductionists*. Causal reductionists are opposed on two sides. On one side are those who deny that there are any high-level causal relations to be reduced—those who think that the causal relations between fundamental physical states of the world are all the causal relations that there are. Call the members of this camp *causal eliminativists*. On the other side are those who, like causal reductionists, accept the existence of high-level causal relations between cigarettes and lung cancer, carbon emissions and climate change, and monetary policy and inflation; however, unlike the causal reductionists, they deny that these high-level causal relations are reducible to low-level causal relations between fundamental physical events. Call the members of this camp *causal emergentists*.

Michael Strevens has the following to say about causal emergentism:

> Some philosophers suspect that...there are irreducible high-level causal relations...Given what we now know, these suspicions are, I believe, extravagant: there simply are no causal relations of

* Forthcoming in *The Journal of Philosophy*
✉: jdmitrigallow@pitt.edu
† I am indebted to Chloe Armstrong, Dave Baker, Gordon Belot, Billy Dunaway, Allan Gibbard, Jim Joyce, David Kaplan, David Manley, Calvin Normore, Charles Sebens, Jon Shaheen, Brian Weatherson, and two anonymous reviewers for comments on earlier versions of this material. Thanks also to audiences at UCLA, Illinois Wesleyan University, Leeds University, Lingnan University, and Koç University.

[1] Menzies (1988)

> which we are aware that cannot be attributed to to lower-level interaction and, ultimately, to the causal influence of fundamental particle on fundamental particle.[2]

My primary goal here is to persuade you that Strevens is incorrect—that, given what we now know, anything *other* than causal emergentism is extravagant.

In section 1, I will introduce and clarify the division between theories of causality which are causally reductionist, causally eliminativist, and causally emergentist. I will illustrate the division between causal reductionism and causal emergentism with David Lewis's two theories of causation, one of which entails causal emergentism, the other of which entails causal reductionism. I will illustrate causal eliminativism with the views of Michael Strevens. In section 11, I will argue that causal reductionism and causal eliminativism are incapable of adequately accounting for the apparent abundance of causal relations between fine-grained events and the apparent scarcity of causal relations between coarse-grained events. In response, I will counsel a rejection of causal reductionism and causal eliminativism, and an acceptance of causal emergentism. As I will explain in more depth below, causal emergentism is consistent with the thesis of *microphysical reductionism*—the thesis that all facts reduce to microphysical facts. In order to be causal emergentists, we need only deny that high-level causal facts reduce to certain microphysical *causal* facts, and not that they reduce to *any* microphysical facts.

1 The Division

In 1973, David Lewis outlined a counterfactual theory of singular causation.[3] According to this theory, an event e causally depends upon a distinct[4] event c iff, had c not occurred, e would not have occurred either,

$$e \textit{ causally depends upon } c \iff \neg O(c) \boxdot\!\!\rightarrow \neg O(e)$$

This counterfactual is evaluated according to Lewis's semantics:[5] '$\neg O(c) \boxdot\!\!\rightarrow \neg O(e)$' is true iff the closest world in which c fails to occur is a world in

[2] Strevens (2008, p. 82) When Strevens says that high-level causal relations are "reducible", he means that they are reducible to low-level causal relations.

[3] Lewis (1973a)

[4] Here, 'distinct' means a bit more than 'non-identical'. It must also be the case that e does not *imply*, and is not implied by, c and that c and e not overlap. See Lewis (1986).

[5] See Lewis (1973b, 1979)

which e fails to occur as well.[6] Causation is the ancestral, or the transitive closure, of causal dependence. That is, one event c causes another event e iff there is a chain of events running from c to e such that, for each link in the chain, the second event causally depends upon the first.

Because the 1973 account determines whether e causally depends upon c by looking to the closest world at which c fails to occur, it requires a sharp line to be drawn between those possibilities in which an event occurs in a different *manner* and those possibilities in which the event simply fails to occur at all. For instance, suppose that in the actual world, Suzy throws the rock overhand. In order to evaluate whether this throw caused the bottle to shatter, Lewis's 1973 account is forced to legislate on whether a world in which Suzy throws the rock underhand is a world in which Suzy's actual throw occurs, but in a different manner, or whether it is a world in which Suzy's actual throw is replaced by a different throw. In his 2000 revision of the counterfactual account,[7] Lewis expresses doubt about the prospects for drawing this line in general. He thus abandons the strategy of looking to the closest world at which the cause fails to occur, and instead focuses on *all* the possible alterations of the time, manner, or occurrence of an event, whether those alterations constitute the same event or not. Whether Suzy's throwing underhand is a different event from her actual throw or not, it will still constitute an *alteration* of her actual throw. Lewis then says that an event c *causally influences* a distinct event e iff there is a substantial range of not-too-distant alterations of c, $c_1 ... c_N$, and a range of alterations of e, $e_1 ... e_N$ (at least some of which differ), such that the alterations of c counterfactually pattern with the alterations of e. That is, c causally influences e iff, for all i, had c_i occurred, e_i would have occurred.

$$c \textit{ causally influences } e \iff O(c_1) \boxdot\!\!\rightarrow O(e_1) \wedge \ldots \wedge O(c_N) \boxdot\!\!\rightarrow O(e_N)$$

Causation is the ancestral of causal influence. That is, c causes e iff there is a chain of events running from c to e such that, for each link in the chain, the first event causally influences the second.

On the surface, these two accounts appear very similar. In Lewis's words, we've simply traded whether-upon-whether counterfactual dependence for whether-, when-, and how-upon-whether, -when, and -how counterfactual dependence. Despite their similarities, there is one dimension along which

[6] If we drop the assumption that there is such a world, then '$\neg O(c) \boxdot\!\!\rightarrow \neg O(e)$' is true just in case there is a possible world in which neither e nor c occurs which is closer than any world in which e occurs and c does not. The standards of similarity are those specified in Lewis (1979).

[7] Lewis (2000)

the theories differ notably. Lewis's 2000 account entails *causal reductionism*. His 1973 account, on the other hand, entails *causal emergentism*. Allow me to explain.

1.1 Causal Reductionism

Some prefatory remarks: in what follows, I will call the entity denoted by a nominalization flanking the verb 'cause' in a causal claim like 'Chris's smoking caused his contraction of cancer' an *event*. I use this word simply because most theories of causation claim that it is events, as opposed to facts, which are causally related. However, I mean to leave it open what the causal relata are. I'll call an event picked out by a nominalization describing a region of spacetime in maximally specific detail with just the predicates of fundamental physics a *finely-individuated*, or *low-level*, event. Low-level events pin down precisely what is happening, at a fundamental physical level, within a given region of spacetime. Events picked out by nominalizations involving predicates more coarse-grained than this ('chemical reaction', 'birthday party', or 'economic depression', *e.g.*) I'll call *coarsely-individuated* events. If an event is distinct from every low-level event, then I'll call it a *high-level* event. So, if you think that stagflation is just the event of thus-and-such fundamental entities being arranged thus-and-so with thus-and-such fundamental properties over a particular time period, then, even though stagflation is coarsely-individuated, you do not think that it is a high-level event, as I am using that term. That is: I leave it open whether coarsely-individuated events are high- or low-level events. I will assume throughout, however, that if there are high-level events, then most coarsely-individuated events are high-level. Note also that, while disjunctions of nominalizations picking out low-level events utilize only the predicates of fundamental physics, they do not describe a region of spacetime in maximally specific detail; so, the event picked out by such a disjunction will not count as a low-level event unless it is identical to an event picked out by one of its disjuncts. Indeed, one may think that every high-level event is just a (perhaps infinitary) disjunction of low-level events. Instances of the causal relation between high-level events, I'll call *high-level causal relations*, and instances of the causal relation between low-level events, I'll call *low-level causal relations*.

Jaegwon Kim gives voice to the position I'm calling *causal reductionism* when he writes that

> macro-causality...must be viewed not as something basic and fundamental but as something that is reducible to, and explain-

> able in terms of, more fundamental causal processes.[8]

On Kim's view, the reduction of macro-causality to micro-causality is to be carried out as follows:

> if the macrocausal relation to be reduced is one from an instance of a property F to an instance of property G, we need to correlate F with some micro-property f, and also G with g, and then we show that f and g are appropriately causally connected.[9]

For Kim, an event is just a property exemplification, so when he talks about macro- and micro-properties, he is talking about the macro- and micro-*events* of those properties being exemplified.[10] The kind of correlation between the macro-properties and the micro-properties that Kim has in mind is the relation of supervenience. To correlate F with f is to show that F supervenes upon f. However, for the purposes of characterizing the position I'm going to call *causal reductionism*, we can remain neutral both on the metaphysics of events and on what relation it is that the relevant low-level events bear to the relevant high-level events. Let's just call that relation, whatever it is, 'realization'. For Kim, e realizes E iff E's property supervenes upon e's property; another causal reductionist might think that e realizes E iff E's occurrence is grounded in e's occurrence; and, of course, there are other positions available. Kim's view, then, is that, as a matter of nomic necessity, when c realizes C and e realizes E, C's causing E reduces to c's causing e. (Throughout, I'll use the uppercase 'C' and 'E' as variables ranging over high-level events, and I'll use the lowercase 'c' and 'e' as variables ranging over C's and E's respective low-level realizers.)

If, as a matter of nomic necessity, A reduces to B, then the material conditional $A \Leftrightarrow B$ will be nomically necessary, so the above thesis entails

> CAUSAL REDUCTIONISM[11]
> Of nomic necessity, where c realizes C and e realizes E,
>
> $$C \textit{ caused } E \iff c \textit{ caused } e$$

[8] KIM (1984b)

[9] KIM (1984a)

[10] See KIM (1976)

[11] CAUSAL REDUCTIONISM should be understood in such a way that it is false if C and E have no low-level realizers. That is, the logical form of CAUSAL REDUCTIONISM is:

$$\Box\ \forall C, E\ [\textsc{cause}(C, E) \Leftrightarrow\ \exists c, e\ [[r(C) = c \wedge r(E) = e] \wedge \textsc{cause}(c, e)]]$$

where r is a function taking high-level events to their low-level realizers.

For the purposes of this paper, I'm going to take this weaker thesis—along with the assumption that some high-level events are causally related—to define the thesis I'm calling *causal reductionism*.[12]

We might worry that not all high-level events are realized by some low-level event. It could be that negative high-level events occur in virtue of the fact that *no* low-level event of the appropriate kind occurs, for instance. Or it could be that which high-level event occurs at a time and place depends upon more than the low-level state of the world at that time and place. For better or worse, the causal reductionist is committed to the claim that every high-level event which enters into causal relations has some low-level realizer. Without this assumption, they cannot even formulate their view. Causal reductionism should, therefore, be understood as entailing that there is a realization relation between some low-level event and every causally-related high-level event. If this claim is false, then causal reductionism is false (see note 11). In contrast, neither the causal eliminativist nor the causal emergentist is committed to the claim that any high-level events are realized by low-level events.

Some self-described causal reductionists might wish to formulate their view slightly differently than Kim. For instance, Huw Price writes that

> many of us think that there is something fundamental about microphysics...As a result, we are attracted to the idea that macroscopic causation is constituted by a lot of microscopic causation. One aspect of this intuition is that causal connections...decompose 'vertically' into a *complex* of microphysical causal relations.[13]

Price proposes that high-level events are realized, not by a *single* low-level event, but rather *a complex* of low-level events, and that what it is for two high-level events to be causally related is just for a complex of causal relations to obtain between the low-level events which realize them. How many of the low-level events realizing C must be causally related to the low-level events realizing E? And how many of the low-level events realizing E must they be related to? Is it enough that a single low-level event among those realizing C cause a single low-level event among those realizing E? Difficult questions, all. It would be better if Price could avoid them entirely. Perhaps he can. Price tells us that each high-level event is realized by a complex of

[12] The reason for this is that I am interested in divisions between philosophical theories of causation, and even though many extant theories of causation entail either causal reductionism or its negation, those which entail causal reductionism are not usually up front about whether they take C's causing E to *reduce* to c's causing e or merely be nomically necessitated by c's causing e.

[13] Price (1992)

many low-level events. Take their mereological fusion.[14] Then, Price could simply say that two high-level events are causally related if the fusions of the complex of low-level events realizing them are causally related. Such a view would be equivalent to the thesis I called *causal reductionism* above. In any case, whether we accept the reformulation or not, the argument I will provide in section II will apply, *mutatis mutandis*, to Price's version of causal reductionism as well. So it will not matter, at the end of the day, whether I can squeeze Price into the Kimian framework. A Price-style reductionist will run into precisely the same troubles.

1.2 Causal Eliminativism

Causal reductionists accept that there *are* high-level causal relations, distinct from the low-level causal relations into which they decompose. They accept that Chris's smoking caused his cancer and that Zimbabwe's monetary policy caused its hyperinflation. They simply believe that these causal relations can be reduced to causal relations between low-level events. There is another position which denies that there are any high-level causal relations to be reduced. Call this position *causal eliminativism*. In general, causal eliminativists believe that the causal relations between low-level events are all the causal relations there are.[15]

There are two ways of being a causal eliminativist. Firstly, you could deny that there are any high-level events. That is, you could think that 'The Weimar Republic's monetary policy' and 'the Mark's hyperinflation' denote low-level events. Then, if you accept the truth of the causal claim 'The Weimar Republic's monetary policy caused the Mark's hyperinflation', you take the causal relation asserted by this claim to be a relation between two low-level events. Call this position *strong causal eliminativism*—strong because it does away with both high-level causal relations and high-level events.

Secondly, you could retain high-level events, and simply deny that these high-level events are causally related. Call this position *weak causal eliminativism*—weak because it only does away with high-level causal relations, retaining the high-level events. It may be thought that weak causal elimina-

[14] I'm presupposing a theory of events according to which they are just regions of spacetime, or properties of regions of spacetime. If you think that events are facts, take the conjunction of these facts.

[15] Assuming that they think that there are low-level causal relations. Denying the existence of both high- and low-level causation is another way of being a causal eliminativist. Russell (1912) provides an example of this brand of causal eliminativism. As I point out in the body with respect to weak causal eliminativism, without something like the Davidson-Strevens thesis, this position entails that 'carbon emissions caused climate change' is unassertible, and is, I take it, therefore effectively refuted by the IPCC report.

tivism provides its own *reductio*. To accept that there are high-level events yet deny that these events enter into causal relations is to deny that carbon emissions caused global temperatures to rise. It is to deny that asbestos ever caused mesothelioma and that oxidation ever caused rust. Intelligent people ought not deny such things, and certainly not on account of philosophical considerations alone. Unfortunately, weak causal eliminativism cannot be dismissed so quickly. Sophisticated weak eliminativists will accept the assertibility of the English sentence 'Chris's smoking caused his cancer', while denying that this sentence is used to assert the existence of a causal relation between the smoking and the cancer. The most developed version of this account that I am aware of comes from Michael Strevens. According to Strevens,

> claims of the form *c was a cause of e*...do not assert the existence of a raw metaphysical causal relation between two events *c* and *e*; rather, they are causal-explanatory claims that assert that *c* is a part of the causal explanation for *e*.[16]

Strevens is here picking up on a suggestion of Davidson's: namely that, in sentences like 'The collapse was caused, not by the fact that the bolt gave way, but rather by the fact that it gave way so suddenly and unexpectedly', the verb 'caused'

> is not the 'caused' of straightforward singular causal statements, but is best expressed by the words 'causally explains.'[17]

What Davidson suggests as a fix for a few recalcitrant sentences, Strevens adopts for all, or at least most, of our everyday causal claims. I'll call this thesis, that causal claims are causal explanatory claims, the 'Davidson-Strevens thesis'.

There are two ways of understanding the Davidson-Strevens thesis. We could understand it as a semantic claim: the literal content of '*C* caused *E*' is that *C* is a part of a causal explanation of *E*. Alternatively, we could understand it as a claim about the pragmatics of causal claims: in ordinary contexts at least, to say '*C* caused *E*' is to pragmatically implicate that the low-level realizer of *C* is causally related to the low-level realizer of *E* and that the occurrence of *C* explains the occurrence of *E*, even though the literal content of '*C* caused *E*' is just that *C* and *E* are causally related. This pragmatic story, wedded with weak causal eliminativism, entails that most causal claims are used to implicate true things despite being strictly speaking false. (This is not as odd as it may seem. On many semantic theories, we often use false

[16] Strevens (2008, p. 4, see also ch. 6)
[17] Davidson (1967, pp. 161–2)

sentences to pragmatically implicate something true—*e.g.*, 'You cannot get there from here' and 'I have nothing to wear.')

This sophisticated form of weak causal eliminativism is, therefore, not merely a metaphysical claim; it is also a claim about the semantics (or the pragmatics) of causal talk. Of course, the view that causal claims are (or pragmatically implicate) causal explanatory claims could be combined with any of the positions I am considering here. However, for the weak causal eliminativist, this thesis, or something like it, is necessary to avoid the absurd conclusion that the sentence 'carbon emissions caused climate change' is unassertable.

1.3 Causal Emergentism

The remaining logical space is occupied by a character I will dub the *causal emergentist*. Many of the so-called British emergentists embraced emergentism about laws of nature. They held that the higher-level laws of nature governing complex phenomena such as chemical reactions or the biological activity of living organisms did not reduce to the fundamental laws of physics.[18] To have a name, call this kind of emergentism *nomic emergentism*. Just as the nomic emergentist thinks that there are higher-level laws of nature which cannot be reduced to lower-level laws of nature, the *causal emergenist* believes that there are causal relations between high-level events which cannot be reduced to causal relations between their low-level realizers.[19] They agree with the causal reductionist, against the causal eliminativist, that there is high-level causation. However, they reject the biconditional

$$C \textit{ caused } E \iff c \textit{ caused } e$$

(where c and e are the low-level realizers of C and E, respectively) for some C and E at some nomologically possible world.

Note that, in rejecting this biconditional, the causal emergentist need not commit themselves to the claim that there *are* low-level realizers of high-level events. Rejecting the idea that every causally-related high-level event is realized by some low-level event is just one way of being a causal emergentist. Granting that every causally related high-level event has some low-level re-

[18] See McLaughlin (1992)

[19] It is for this reason, and this reason alone, that I choose the name 'causal emergentism'. The position is considerably weaker than many of the claims advanced by philosophers calling themselves 'emergentists'. The weakness of the position is, to my mind, one of the position's strengths—it allows us to account for the sense in which the high-level causal facts swing free of the low-level causal facts without making any of the more metaphysically weighty claims typically advanced by emergentists.

alizer, there are at least two different ways to reject this biconditional. On the one hand, we could say that there are some high-level events which are not causally related, even though their low-level realizers are. Alternatively, we could say that there are causally-related high-level events whose low-level realizers are not causally related. Either of these positions is consistent with *causal emergentism*. However, given the kinds of laws that exist at our world, I believe that any arbitrary pair of non-simultaneous low-level events,[20] e_1 and e_2, will be causally related to one other—*i.e.*, either e_1 caused e_2 or e_2 caused e_1 (this is for the reasons discussed in section 11 below). Therefore, I find the first way of being a causal emergentist far more plausible than the second.

Causal emergentism does not entail nomic emergentism. Moreover, you could be a causal emergentist without denying that causal relations between high-level events can be reduced to *some* low-level facts. In order to count as a causal emergentist, it is enough to claim that the low-level facts in virtue of which high-level events are causally related are not exclusively low-level *causal* facts about the realizers of those high-level events. You *may*, of course, accept a higher-octane version of emergentism, according to which causal relations between high-level events are not determined by or in principle predictable on the basis of the low-level state of the world and the low-level laws of nature; but there is nothing in the thesis of causal emergentism demanding such a claim. Causal emergentists can accept that the behavior of every physical object is entirely determined by the fundamental physical state of the world and the fundamental physical laws.[21]

To see this, return to Lewis's 1973 account of causation. Lewis accepted that the high-level causal facts were determined by the low-level facts.[22] Nevertheless, his original counterfactual account of causation entails causal emergentism. On that account, recall, an event e causally depends upon a distinct event c iff were c to fail to occur, e would fail to occur as well. Causation is the ancestral of causal dependence. This account is in need of a theory of events, and the conditions under which they do and do not occur. This was provided by Lewis's 1986 paper *Events*.[23] There, Lewis claims that an event is

[20] In light of special relativity: any arbitrary pair of time-like separated low-level events

[21] Along, perhaps, with some brute facts about the outcome of tychistic chancy processes, like the collapse of the wave function on some interpretations of quantum mechanics.

[22] Given Lewis's physicalism, the low-level facts will specify a unique possible world. The laws at this world come along for free—they are the generalizations of the axiomatic system of truths which strikes the best balance of simplicity, strength, and fit (see Lewis, 1983, 1994) The semantics for counterfactuals (see Lewis, 1979) then provides the truth conditions for causal relations both high and low.

[23] Lewis (1986)

a property of a spacetime region. Since, for Lewis, properties are just classes of individuals at worlds, a property of a spacetime region is just a class of spacetime regions at worlds. A spacetime region at a world has the property just in case it belongs to the class. For an event e to occur at a world is for one of its members to exist at that world.

Lewis tells us that some events *imply* other events, where an event e implies an event f iff, necessarily, if e occurs in region R, then f occurs in region R. He illustrates this kind of implication with the following example: John's saying 'hello' loudly implies John's saying 'hello'. Though he expresses some hesitation at taking these two events to be distinct, Lewis decides that they must differ, since they differ causally. John's saying 'hello' caused Fred to greet him, whereas John's saying 'hello' loudly did not. (Fred would still have greeted John even if he had said 'hello' at a normal volume.) Similarly, John's saying 'hello' loudly was caused by John's being a bit drunk, whereas John's saying 'hello' was not caused by John's being a bit drunk. (Had John not been drunk, he would not have said 'hello' loudly; however, he would still have said 'hello'.) Lewis describes the relation between these two events thusly:

> We have, so to speak, a more and a less detailed version of what happens in a region. Both are occurrent events. The more detailed version has a richer essence; the otherworldly regions included in it are fewer and less varied...The more detailed version is one, but only one, of the ways in which the less detailed version could have occurred.[24]

The relationship between high-level events and their low-level realizers is analogous to the relationship between John's saying 'hello' and John's saying 'hello' loudly. Low-level events imply the high-level events they realize. The low-level event is a more detailed version of what happens in a region; it has a richer essense than the high-level event it realizes. The low-level realizer is one, but only one, of the ways in which the high-level event could have occurred. For this reason, just as John's saying 'hello' loudly can be caused by things that John's saying 'hello' is not, a low-level event can be caused by things which the high-level event it realizes is not. That is because the closest possible world at which the cause fails to occur could be a world at which the low-level realizer fails to occur, yet the high-level event it realizes does not fail to occur.

In addition, two low-level events c and e can be causally related without the high-level events they realize, C and E, being causally related. For the

[24] Lewis (1986, p. 257)

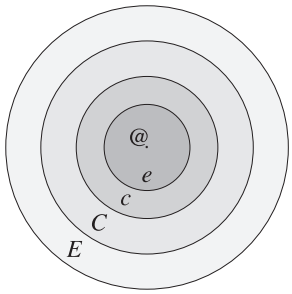


FIGURE 1: In the diagram, simiarity is represented with distance. All and only the worlds inside the innermost circle are worlds at which e occurs; all and only those inside the next largest circle are those at which c occurs; and so on. The closest world to the actual world, @, at which c does not occur is a world at which e does not occur. Yet the closest world at which C does not occur is not a world at which E does not occur.

closest world at which c fails to occur could be a world at which e fails to occur without the closest world at which C fails to occur being a world at which E fails to occur. For instance, suppose that the modal profiles of c, e, C, and E are as shown in figure 1. There, the counterfactual '$\neg O(c) \mathrel{\Box\!\!\rightarrow} \neg O(e)$' is true, while the counterfactual '$\neg O(C) \mathrel{\Box\!\!\rightarrow} \neg O(E)$' is false. Suppose further that there is no sequence of events $D_1 ... D_N$ such that $\neg O(C) \mathrel{\Box\!\!\rightarrow} \neg O(D_1) \wedge ... \wedge \neg O(D_N) \mathrel{\Box\!\!\rightarrow} \neg O(E)$. Then, Lewis's 1973 account will tell us that c caused e even though C did not cause E.

Due to the universality of fundamental physical forces like gravitation and electromagnetism, examples with this structure are not difficult to come by.[25] The moon wanes, and the window shatters. The waning of the moon is realized by the fundamental particles which constitute the moon having certain masses and charges, and taking certain precise trajectories over a certain period of time. Similarly, the shattering of the window is realized by the individual particles which constitute the window having certain masses and charges and taking certain precise trajectories over a certain period of time. Both of these are incredibly fragile events. Because their essences are so rich, it is very easy for them to fail to occur—if any of the particles in the window were to have an ever-so-slightly different mass or charge or take

[25] In the body, I'm going to pretend that the fundamental physical state of the world and the fundamental physical laws of nature are roughly what we thought that they were about a century and a half ago. I do not believe, however, that the arguments crucially depend upon this assumption. Under more realistic assumptions, extra provisos about past light cones and the like will be needed. When appropriate, I'll make the necessary revisions in the footnotes.

an ever-so-slightly different trajectory during the time period, then the low-level realizer of the shattering would fail to occur; and likewise for the masses, charges, and trajectories of the particles in the moon and the low-level realizer of the moon's waning. Thus, had the low-level realizer of the moon's waning failed to occur, at least one of the particles which constitute the moon would have had a slightly different mass or charge or taken a slightly different trajectory. However, if any of those particles had had a slightly different mass or charge or taken a slightly different trajectory, then it would have exerted a slightly different gravitational or electromagnetic force on all of the particles which realize the window's shattering. Since the future trajectories of the particles in the window are entirely determined by the resultant of the forces acting upon them, they would have taken slightly different trajectories, had any of the gravitational or electromagnetic forces acting upon them been slightly different. So the low-level realizer of the window's shattering would have failed to occur. So the low-level realizer of the moon's waning caused the low-level realizer of the window's shattering, on Lewis's 1973 account. But the moon's waning did not cause the window's shattering, since, had the moon not waned, the window would still have shattered. (It would have shattered in an ever-so-slightly different *way*, but it would have shattered all the same.) Assuming that there is no intermediate event which depends upon the moon's waning and upon which the shattering depends, it follows on the 1973 account that the moon's waning did not cause the window's shattering.

So, on Lewis's 1973 counterfactual account of causation, if you look at the low-level state of the world, then you will see one causal structure. If you look at the high-level state of the world, you will see a different causal structure. And which causal structure you will see depends upon which high-level events you're looking at. There is one network of causes and effects leading into and out of John's saying 'hello' and another network of causes and effects leading into and out of John's saying 'hello' loudly.

Interestingly, this aspect of the counterfactual account goes away in Lewis's 2000 revision. On the revised account, recall, an *alteration* of an event is just a variation of the time or manner of the event's occurrence, whether that variation leads to the event failing to occur, or merely occurring in a slightly different manner, or at a slightly different time. An event c causally influences a distinct event e iff there is a substantial range of not-too-distant alterations of c, $c_1 ... c_N$, and a range of alterations of e, $e_1 ... e_N$ (at least some of which differ), such that the c_i counterfactually pattern with the e_i—that is to say, had c_i occurred, e_i would have occurred, for all i. Causation is the ancestral, or the transitive closure, of causal influence.

On this account, no longer can John's saying 'hello' loudly be caused by

something which does not also cause John's saying 'hello'. If not-too-distant alterations of John's being drunk counterfactually pattern with alterations of John's saying 'hello' loudly, then they will also counterfactually pattern with alterations of John's saying 'hello'. That's because an alteration of John's saying 'hello' loudly *just is* an alteration of John's saying 'hello'. A variation in the time or manner of the event of John's saying 'hello' loudly just is a variation in the time or manner of the event of John's saying 'hello.' This variation might be a variation which makes it the case that John's saying 'hello' loudly does not occur, while his saying 'hello' still does occur, but that does not matter, on the 2000 account. So long as alterations in John's being drunk counterfactually pattern with alterations of John's saying 'hello', it does not matter whether those alterations are ones in which the saying 'hello' occurs in a different manner, or at a different time, or whether they are ones in which it does not occur at all.

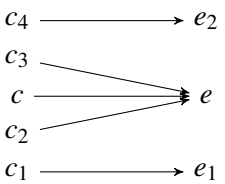
For similar reasons, if not-too-distant alterations of the low-level realizer of the moon's waning counterfactually pattern with alterations of the low-level realizer of the window's shattering, then not-too-distant alterations of the moon's waning will counterfactually pattern with alterations of the window's shattering. That's because not-too-distant alterations of the low-level realizer of the moon's waning *just are* not-too-distant alterations of the moon's waning; and alterations of the low-level realizer of the window's shattering *just are* alterations of the window's shattering. So, if a low-level event c influences another low-level event e, then the high-level event C which c realizes must influence the high-level event E which e realizes—for *any* high-level events C and E which c and e realize. So, if c influences e, then C influences E.[26]

This is just a claim about influence. Causation, however, is the ancestral of influence. If c causes e by influencing it directly, then, since—as we just saw—if c influences e, then C influences E, we can conclude that C influences E, and therefore, that C causes E. If, on the other hand, c causes e by being connected to it by a chain of influence $c \rightarrow d_1 \rightarrow d_2 \rightarrow ... \rightarrow d_N \rightarrow e$, then we know that not-too-distant alterations of c counterfactually pattern with alterations of d_1. But not-too-distant alterations of c just are not-too-distant alterations of C, so not-too-distant alterations of C must also counter-

[26] We might worry about the requirement that the range of alterations of C be *substantial*. It could be that what counts as a substantial range of alterations for c does not count as a substantial range of alterations for C. Perhaps. But even so, the range of alterations of c which counterfactually pattern with alterations of e will be at least a subset of a substantial range of alterations of C; and if a set of alterations counterfactually pattern with a range of alterations of e, then any superset will also counterfactually pattern with a range of alterations of e, since Lewis allows arbitrarily many of the alterations of e to be identical, so long as some of them differ.

factually pattern with alterations of d_1. And, since d_N influences e, not-too-distant alterations of d_N counterfactually pattern with alterations of e. But alterations of e just are alterations of E, so not-too-distant alterations of d_N must also counterfactually pattern with alterations of E. So there is a chain of causal influence running from C to E *via* $d_1, d_2, ..., d_N$. Since causation is the ancestral of causal influence, C caused E. So, whether c caused e by influencing it directly or by being linked to e by a chain of influence, if c caused e, then C caused E. So Lewis's 2000 account entails the right-to-left direction of CAUSAL REDUCTIONISM.

Were it not for Lewis's restriction to "not-too-distant" alterations, we could run the very same argument in reverse to establish the left-to-right direction as well. Such an argument would founder on the fact that alterations which are not-too-distant from a high-level event may count as too distant from the low-level realizer of that event. Thus, it may be that, *e.g.*, c_1, c_2, c_3, and c_4 are all alterations of the low-level event c (which realizes the high-level event C); and e_1 and e_2 are alterations of the low-level event e (which realizes the high-level event E). Then, it may be that the alterations of c counterfactually pattern with the alterations of e in the following way:



If c_1 and c_4 are not-too-distant alterations of the high-level event C but too-distant alterations of the low-level event c, then it may be that C influences E even though c does not influence e.

Patterns of counterfactual dependence like this no doubt arise at some possible worlds, but it does not appear that they will arise at worlds with laws like ours. At worlds like ours, the fundamental laws of nature take the form of differential equations specifying how certain fundamental physical properties will change over time as a function of other fundamental physical properties. In these equations, if large changes in the determining properties lead to changes in the determined properties, then smaller changes in the determining properties will lead to changes in the determining properties, too. They will in general lead to *smaller* changes, but they will lead to changes nonetheless (as in the case of the moon's waning and the window's shattering discussed above). Given Lewis's physicalism, all alterations of events are alterations of the fundamental physical properties of the event's realizer. This means that, if large alterations of c (ones which are too-distant

from c, though not-too-distant from C) counterfactually pattern with alterations of e (as they must if C is to influence E), then smaller alterations of c (ones which are not-too-distant from c) will counterfactually pattern with alterations of e as well. They will in general counterfactually pattern with *less distant* alterations of e, but they will counterfactually pattern with alterations of e nonetheless. Thus, patterns of counterfactual dependence like the one depicted above will be ruled out at worlds with laws of nature like ours. So it seems that the 2000 account will entail the left-to-right direction of CAUSAL REDUCTIONISM at worlds with laws like ours.

But place that exegetical question to one side; it is interesting, but unimportant for my purposes. Wherever the 2000 revision of the counterfactual account falls with respect to the causal emergentist/causal reductionist divide, the important point to stress here is that *causal* emergentism, unlike *nomic* emergentism, does not entail that there are irreducible higher-level laws or even that high-level causal relations are irreducible. Causal emergentists can, like Lewis (1973), take high-level causal relations to be entirely reducible to low-level facts without taking them to be entirely reducible to low-level *causal* facts. So when Kim writes that

> macro-causality...must be viewed not as something basic and fundamental but as something that is reducible to, and explainable in terms of, more fundamental causal processes[27]

he poses a false dichotomy. We need not choose between claiming that high-level causal relations are basic and fundamental and claiming that they are reducible to low-level causal relations. We can insist that the high-level causal relations are entirely reducible to low-level facts without claiming that they are entirely reducible to low-level *causal* facts.

1.4 A TAXONOMY

In summary, we can categorize these three positions according to whether they agree with the following claims.

High-Level Causation: There are high-level causal relations.

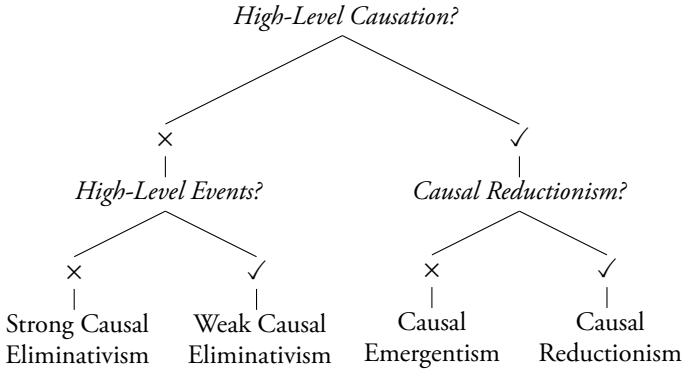
High-Level Events: There are high-level events.

Causal Reductionism: Of nomic necessity, where C is realized by c and E is realized by e, C caused E iff c caused e.[28]

[27] KIM (1984b, p. 51)

[28] Again, we should understand this thesis in such a way that it is false if there is no appropriate realization relation between causally related high- and low-level events. (See note 11.)

If you reject *High-Level Causation*, then you are a causal eliminativist. If you additionally reject *High-Level Events*, then you are a strong causal eliminativist. If you additionally accept *High-Level Events*, then you are a weak causal eliminativist. If you accept *High-Level Causation* and additionally accept *Causal Reductionism*, then you are a causal reductionist. If you accept *High-Level Causation* but deny *Causal Reductionism*, then you are a causal emergentist.



2 A Defense of Causal Emergentism

In this section, I'm going to argue that both causal reductionism and causal eliminativism are untenable. In brief, the problem I will pose for the causal reductionist is that high-level causal relations appear to be relatively sparse, whereas low-level causal relations appear to be relatively abundant. These two claims are in tension with the thesis that high-level events are causally related iff the low-level events which realize them are causally related. The same objection applies, *mutatis mutandis*, to the strong causal eliminativist. The tension between the apparent abundance of low-level causation and the apparent sparsity of high-level causation has been noted before,[29] but it has been thought that the tension can be massaged and mitigated in various ways without going in for causal emergentism. In this section, I'm going to try to make trouble for those mitigating strategies. In the course of stirring up that trouble, I'll provide an independent argument against what I earlier called the Davidson-Strevens thesis—that causal claims are (or implicate) causal explanatory claims. Since weak causal eliminativism is only as plausible as the Davidson-Strevens thesis—without this thesis, the weak eliminativist is forced to say that sentences like 'carbon emissions caused climate change' are

[29] See, *e.g.*, Russell (1912), Loewer (2001), Field (2003), Price & Corry (2007) and Ney (2009).

unassertable—an argument against the Davidson-Strevens thesis is an argument against weak causal eliminativism as well. I will therefore counsel a rejection of causal reductionism and causal eliminativism, and an acceptance of causal emergentism.

More carefully, the problem for causal reductionism and strong causal eliminativism is this: the following three claims form an inconsistent set.

1) A coarsely-individuated event C caused another coarsely-individuated event E if C's low-level realizer caused E's low-level realizer.

2) The orbit of Gliese 163c did not cause stagflation.

3) The low-level realizer of the orbit of Gliese 163c did cause the low-level realizer of stagflation.

(Gliese 163c is a potentially habitable planet, approximately 49 light years from Earth, first discovered in 2012. Throughout, let's take the phrase 'the orbit of Gliese 163c' to refer to Gliese 163c's taking a single trip around its star, beginning at the start of Earth year 1900.) In (1), I'm going to understand the phrase 'low-level realizer' in such a way that, if C is low-level, then C's low-level realizer is itself. Then, the strong causal eliminativists thinks that (1) says that if C caused E, then C caused E. Since this is a tautology, the strong causal eliminativist is committed to (1). The causal reductionist takes (1) to be an entailment of their view, so they are committed to it as well. I will contend that both (2) and (3) are true. Since (2) and (3) are true, (1) is false. From the falsehood of (1), it follows that both causal reductionism and strong causal eliminativism are false.

In support of (3): every major theory of causation in good standing gets the result that the low-level realizer of Gliese 163c's orbit caused the low-level realizer of stagflation. Start with the counterfactual account. As we saw above with respect to the moon's waning and the window's shattering, the low-level realizer of an event like stagflation or the orbit of Gliese 163c consists of certain fundamental physical particles having certain masses and charges and taking certain precise trajectories over a certain period of time. These events are incredibly fragile—if the masses, charges, or trajectories of those fundamental particles were to differ in the slightest, those low-level events would fail to occur. Apply the counterfactual test for causation: were the microphysical realizer of Gliese 163c's orbit to fail to occur, at least one of the particles which realize Gliese 163c's orbit would have differed, however slightly, in its mass, charge, or trajectory; it would therefore have exerted a slightly different gravitational or electromagnetic force upon all of the particles which

realize stagflation. Since the trajectories of those particles are completely determined by the resultant of the forces acting upon them, the trajectories of those particles would have differed, however slightly. So the low-level realizer of stagflation would have failed to occur. So the low-level realizer of stagflation counterfactually depends upon the low-level realizer of Gliese 163c's orbit. No philosopher in the counterfactual tradition claims that non-backtracking counterfactual dependence between distinct events like this is necessary for causation. But almost all accept that non-backtracking counterfactual dependence between distinct events is a sufficient condition for causation.[30] And that is all we need here in order to conclude that the low-level realizer of Gliese 163c's orbit caused the low-level realizer of stagflation.

Consider instead a regularity account of causation. On Mackie's account, for instance,[31] c is a cause of e iff c is a part of a minimally sufficient condition for e that actually obtains. And the low-level realizer of Gliese 163c's orbit, together with the simultaneous state of the rest of the universe, is a minimally sufficient condition for the low-level realizer of stagflation. Given determinism, the entire state of the universe at a time, together with the laws of nature, is sufficient for the state of universe at every other moment—and, in particular, for the part of the universe which realizes stagflation.[32]

Or consider a probabilistic account of causation. According to the most plausible versions of those accounts, c caused e just in case c changes the probability of e in a causally homogenous background context. A causally homogenous background context is given by all of the causes of e, except c

[30] Besides the non-backtracking and distinct event qualifications, we should also add that the counterfactuals relate intrinsic properties of the events in question. If we are allowed to appeal to mere Cambridge properties, then the counterfactual test would imply that a distant supernova could cause me to become such that a supernova has recently occurred. However, many find this result unpalatable (not the least because it seems to involve action-at-a-distance). See (Maudlin, 2007, ch. 5).

[31] Mackie (1965)

[32] On certain interpretations of quantum mechanics, the entire state of the universe at one time only determines a probability distribution over the state of the universe at future times. In that case, we could say that the entire state of the universe at one time, together with facts about the outcome of tychistic chancy processes—the collapse of the wavefunction—constitutes a minimally sufficient condition for the state of the universe at future times; it would then follow that the low-level realizer of Gliese 163c's orbit caused the low-level realizer of stagflation.

Some regularity theorists think that events which had some chance of not occurring at t are incapable of being caused by events occurring at t. Such theorists will want to deny that the low-level realizer of Gliese 163c's orbit caused the low-level realizer of stagflation. Because this position has fallen out of fashion—in part because it precludes quantum mechanical causation—I'm happy to let it fall by the wayside in my argument here. (Thanks to an anonymous referee for pressing me on this point.)

and events caused by c (if c is indeed a cause of e). So, given the universality of the fundamental forces, the entire state of the rest of the universe at a time, other than the realizer of the orbit of Gliese 163c, constitutes a causally homogenous background context for the microphysical realizer of stagflation—call that background context 'K'. K does not entail that the microphysical realizer of stagflation occurs, so, given some plausible assumptions, the probability of the microphysical realizer of stagflation, given K, should be less than 1. However, assuming determinism, the probability of the microphysical realizer of stagflation, given K and the realizer of the orbit of Gliese 163c, will be 1. So, on the probabilistic account, the realizer of the orbit of Gliese 163c caused the realizer of stagflation.[33]

Or consider a process theory of causality, like that of Dowe[34] or Salmon.[35] On Dowe's account, a low-level event c caused another low-level event e if, roughly, c and e are connected by a series of causal processes whose intersections constitute causal interactions. A causal process is just the world line of an object which possesses a conserved quantity, and a causal interaction is an intersection of world lines that involves the exchange of a conserved quantity. Photons collide with the particles which make up the realizer of Gliese 163c's orbit. These collisions constitute causal interactions; in the collisions, there is an exchange of momentum between the particles and the photons. Some of these photons make their way to Earth; some of those collide with some of the particles which make up the low-level realizer of stagflation. These collisions also count as causal interactions; momentum is exchanged between the photons and the particles. So there is a causal process leading from the low-level realizer of Gliese 163c's orbit to the low-level realizer of stagflation. So, on Dowe's process theory, the low-level realizer of Gliese 163c's orbit caused the low-level realizer of stagflation. (Similar remarks apply to Salmon's process theory.)

And since an in-principle intervention upon the state of the low-level realizer of Gliese 163c's orbit would bring about a change in the state of the low-level realizer of stagflation (for the very reason that the counterfactual

[33] Suppose that an indeterministic interpretation of quantum mechanics is correct. Then, the probability of the realizer of stagflation, given the realizer of the orbit of Gliese 163c and K, need not be 1. However, it will still be the case that changes in the quantum state of any of the particles in the realizer of Gliese 163c's orbit will make a difference to the universe's Hamiltonian, which will make a difference to the future evolution of the universal wave function ψ, *via* the Schrödinger equation, which will make a difference to the probability distribution over the low-level state of the universe realizing stagflation. So the low-level state of the orbit of Gliese 163c will change the probability of the low-level realizer of stagflation. (Thanks to an anonymous referee for pressing me on this point.)

[34] Dowe (2000)

[35] Salmon (1994)

'had the low-level realizer of Gliese 163c's orbit not occurred, the low-level realizer of stagflation would not have occurred' comes out true) the manipulationist account of Woodward[36] will rule the realizer of Gliese 163c's orbit a cause of the realizer of stagflation.

Every major philosophical theory of causality in good standing implies that the low-level realizer of Gliese 163c's orbit and the low-level realizer of stagflation are causally related. I conclude that they are causally related. So (3) is true. If (2) is also true, causal reductionism and strong causal eliminativism are both false.

Some respond to these kinds of considerations by rejecting (2). They accept (3), and this, together with their reductionism or strong eliminativism, entails that the orbit of Gliese 163c caused stagflation. So they countenance the counterintuitive causal relationship, but insist that they can explain away the intuitive appeal of (2) with various semantic or pragmatic theses about our causal talk. Here is Alyssa Ney endorsing this general strategy:

> Although it is true that the foundationalist picture of causation under consideration[37] has the consequence that due to the multitude of fundamental causal interactions, there are many, many factors of causal influence for every event we might consider, we ordinarily want to single out at most a few as 'the causes' of an event, in our ordinary causal assertions.[38]

For the remainder of the essay, therefore, I will be considering the plausible semantic or pragmatic theses which it has been supposed are capable of successfully explaining away the intuitive appeal of (2), and arguing that none of them meet with success. Of course, it is worth emphasizing that there will *always* be some pragmatic story the reductionist or eliminativist can tell which will be capable of getting all the data right. Suppose, for instance, that you have produced an emergentist theory of causation, call it 'T_E', which aligns perfectly with our characteristic causal judgments in every case; it says that two events are causally related when and only when we judge the corresponding causal claim to be true. Suppose then that a reductionist puts forward a theory of causation, call it 'T_R', which does a much worse job aligning with our characteristic causal judgments; it frequently asserts the existence of a causal relation when we judge there to be none. Such a reductionist could

[36] Woodward (2003)

[37] While she does not say enough for me to precisely locate her in the taxonomy of section 1.4, the view of causation Ney calls 'foundationalist' is at least committed to the disjunction of causal reductionism and causal eliminativism.

[38] Ney (2009, p. 741)

always put forward the following semantic/pragmatic thesis: a sentence '*c* caused *e*' is true/assertible iff T_E claims that there is a causal relation between *c* and *e*. This will be capable of capturing all of the data iff T_E was capable of capturing all of the data. But, I take it, such an account wears its implausibility on its sleeve. It is not enough to provide some *ad hoc* pragmatic theory of which causal claims are true/assertible. It must additionally *make sense*, given the theory of what the causal relation is and given general features of language, that we would talk in accord with the semantic/pragmatic theory. The two theses I will consider below—*viz.*, that we choose to ignore the causes which have negligible influence, and that we choose to ignore the causes which do not provide good explanations—both meet this minimal standard of adequacy. It is a general feature of language that we neglect the negligible. And, given that we often look to causal relations in order to provide explanations, it makes good sense that we would only focus on those that succeed in providing such explanations.

Turning now to those theses: in the first place, a causal reductionist or a strong causal eliminativist may wish to suggest that, while Gliese 163c's orbit *did* cause stagflation, the causal influence that Gliese 163c's orbit had on stagflation was so minute that, for all practical purposes, it can be ignored. For this reason, we hesitate to accept (2). For instance, Lewis recognizes that his 2000 account commits him to saying that

> ...almost everything that precedes an event will be counted among its causes. By the law of universal gravitation, a distant planet makes some minute difference to the trajectory of Suzy's rock, thereby making a tiny difference to the shattering of the bottle...we open the gate to a flood of spurious causes.[39]

His response to this worry is that we are justified in ignoring these so-called spurious causeson the grounds that their influence will be negligible:

> Well—these differences made by spurious causes are negligible, so surely we are entitled to neglect them.[40]

Here, Lewis diagnoses the oddity of (2) by appeal to a pragmatic thesis about which causes are most felicitously cited in a causal claim. The thesis, which I'll call 'the Lewis thesis', is that the causes with the most influence are most felicitously cited in a causal claim.

---

[39] Lewis (2000, p. 188)

[40] Lewis (2000, p. 189) Interestingly, the quoted sentence changes between the 2000 *Journal of Philosophy* version and the 2004 version in *Causation and Counterfactuals*. While nothing else in the surrounding text changes, this sentence, which ends with a period in the 2000 version, ends with a question mark in the 2004 version.

The Lewis thesis tells us that (2) appears true because, while the orbit of Gliese 163c *did* cause stagflation, in ordinary contexts, we ignore influences as negligible as the orbit of Gliese 163c's—just as, for instance, in ordinary contexts, the sentence 'there is nothing in the fridge' is assertible, even though the sentence is false so long as there is air and dust in the fridge.

Now, I do not think that the right thing to say about Gliese 163c's orbit is that it makes a *negligible* difference to stagflation. Given that not a single job would be saved, not a single firm's production would be higher, and not a single price would be lower without Gliese 163c's orbit—and given that these factors completely determine the precise duration and severity of stagflation—the right thing to say is that Gliese 163c's orbit did not make *any* difference to stagflation, however negligible. But put that point aside. There's a bigger worry. The worry is that, often enough, seemingly spurious causes have quite a large influence on their putative effects—even larger influence than the apparently genuine causes.

Suppose that Sabeen tells you truthfully that she plans to slip a fatal poison into Stephanie's drink. You are unable to warn Stephanie, and you do not know how to neutralize the poison, but you do have on you a powerful anesthetic which will numb and immobilize Stephanie, making her death far less painful. You pour the anesthetic into Stephanie's drink. She drinks and dies quickly and painlessly. According to Lewis's 2000 account, your pouring the anesthetic into Stephanie's drink caused Stephanie's death—since not-too-distant alterations of your pouring of the anesthetic counterfactually pattern with alterations in Stephanie's death; had you not poured, the death would have occurred in a different manner. However, the influence that it had on the death is by no means negligible. Had you not given Stephanie the anesthetic, she would have died a much more painful death. There would have been writhing and cursing and gnashing of teeth. Moreover, the influence of the fatal poison is comparatively small. Given the presence of the immobilizing anesthetic, alterations in the pouring of the poison counterfactually pattern with comparatively minor alterations of Stephanie's breathing, heartbeat, and other metabolic functions. (In fact, we can stipulate that, if the lethal poison had not killed Stephanie, then the anesthetic *would* have eventually prevented her from breathing, causing her to die shortly thereafter. Then, not pouring the poison would only slightly delay the death.) Nevertheless, we are loath to countenance your pouring of the anesthetic into Stephanie's drink as a cause of her death. We judge the sentence

4) # Your pouring anesthetic into Stephanie's drink caused her to die.

to be in some deep sense inappropriate; and we judge the sentence

5) Sabeen's pouring the lethal poison caused Stephanie to die.

to be in some deep sense appropriate. It is my view that (4) is not only inappropriate, but false, but everybody should be able to agree that it is infelicitous to utter (4). The problem is that the Lewis thesis predicts that (4) should be felicitous, or at least *as* felicitous as (5) (if not more so). This prediction is not borne out; which gives us strong reason to reject the Lewis thesis.[41]

In the foregoing argument, I supposed that, if minor alterations of an event c_1 counterfactually pattern with large macroscopic alterations in an event e, and alterations of an event c_2 counterfactually pattern with macroscopically undetectable alterations in e, then c_1 influences e more than c_2 does. In conversation, several people have suggested that a causal reductionist or a strong causal eliminativist should deny this principle, and instead maintain that which influences are greater than others should be a context-sensitive matter. In the case presented above, for instance, they might suggest that, while the poison did not make any macroscopically detectable difference to the event of Stephanie's death; it still made a difference to *whether it was a death*. Such a strategy might help in the present case, but it would fail in general. So understood, the Lewis thesis falls immediately to cases of preemption; alterations of neither Billy's nor Suzy's throw counterfactually pattern with alterations of the window's shattering in which the window does not shatter. So, on the revised account, neither will be appropriately cited as a cause, so long as there is some other event which does make a difference to whether the window shatters—as surely there is: witness the window's being fragile, the shutters' being open, *etc*.

Perhaps the reductionist and the strong eliminativist could borrow a page from the weak eliminativist's playbook, and say that claims of the form '*C* caused *E*' are either causal explanatory claims in disguise, or else pragmatically implicate that *C* causally explains *E*—what I called earlier the 'Davidson-Strevens thesis.' Unfortunately, that thesis does not pan out. In both its semantic and pragmatic flavors, it leads us into bad predictions. Moreover, even if we accept the pragmatic flavor of the thesis, it does not help to explain the infelicity of (2).

In the first place, *C*'s being a part of an adequate explanation of *E* is not necessary for the truth (or the felicity—depending upon which version of the Davidson-Strevens thesis we are arguing against) of the causal claim '*C* caused *E*.' Suppose that you come to me with a bad case of insomnia and ask for something to help you rest better. I hand you a herb and tell you to eat it. You do, and shortly thereafter become sleepy. Suppose that it is common

[41] Similar points are made by Schaffer (2001) and Strevens (2003).

knowledge between us, in this context, that the herb brought about your sleepiness. If you then ask me (6),

6) Why did I get sleepy?

most philosophers of explanation, including Strevens, maintain that (7),

7) Because the herb has a dormitive virtue.

would be a poor answer. In this context, it is common knowledge between us that the herb brought about your sleepiness, and all that it is to have a dormitive virtue is to produce sleepiness. Therefore, in this context, the fact that the herb has a dormitive virtue provides little genuinely explanatory information about why you became sleepy. In this context, there is no adequate explanation of your sleepiness that makes reference to the herb's dormitive virtue. So (8)

8) Your sleepiness is explained by the herb's having a dormitive virtue.

is false. Nevertheless, the causal claim (9)

9) Your sleepiness was caused by the herb's having a dormitive virtue.

appears to be true, or at least felicitous, in this context. Even though citing the fact that the herb has a dormitive virtue does not adequately explain your sleepiness, the herb's having a dormitive virtue still *did cause* your sleepiness. So the adequacy of an explanation whose explanandum is *E* and whose explanans include *C* is not necessary for the truth, or the felicity, of a causal claim '*C* caused *E*'.

Moreover, note that, even if the pragmatic version of the Davidson-Strevens thesis were correct, it would not be sufficient to explain the intuitive truth of (2). That is because, according to the pragmatic version of the Davidson-Strevens thesis, a causal claim '*C* caused *E*' implicates that *C* explains *E*. However, negating a claim which implicates that *p* need not implicate that not-*p*. 'Some of the boys went to the lake' implicates that not all of the boys went to the lake. However, 'None of the boys went to the lake' certainly does not implicate that all of the boys went to the lake. So, even if 'The orbit of Gliese 163c caused stagflation' implicates that the orbit of Gliese 163c is explanatorily relevant to the economy's stagflating, this does not show that 'The orbit of Gliese 163c did not cause stagflation' implicates that the orbit of Gliese 163c *is not* explanatorily relevant to the economy's stagflating. Perhaps if it were obvious that Gliese 163c's orbit *did* cause stagflation, then we could tell some story about how a claim like (2) could be used to implicate

that Gliese 163c's orbit does not *explain* stagflation—*e.g.*, uttering (2) flouts the maxim of quality. However, it is *far* from obvious that Gliese 163c's orbit caused stagflation (in fact, I'm fairly certain that it did not; and I hope to have at least cast the reader into some doubt on this question). And given this, there is no reason to rule out that an utterance of (2) is intended to deny the existence of a *causal* relation—and not merely an *explanatory* relation—between the orbit of Gliese 163c and stagflation.

To sum up: Lewis's strategy for denying (2)—we mistakenly judge (2) to be true because we usually neglect causes whose influence is negligible—founders because high-level events can have substantial and non-negligible effects upon the *way* in which another high-level event occurs, without intuitively causing that high-level event (as demonstrated by the infelicity of (4)). The Davidson-Strevens strategy for denying (2)—claims like (2) are either true in virtue of the fact that the cause does not explain the effect, or else they are false, but mistaken for true because they implicate the true claim that the cause does not explain the effect—founders because, firstly, causal explanatoriness is not necessary for either the apparent truth or the felicity of a causal claim, and secondly, in its pragmatic form, the thesis fails to explain the apparent truth of (2), since denying a claim which implicates that p need not implicate that not-p. More generally, given that it is far from obvious that there is a causal relation between Gliese 163c's orbit and stagflation, there is no reason for a listener to rule out that an utterance of (2) is meant to deny the existence of a causal relation between these two events.

3 In Summation

The straightforward strategies for explaining (2)'s apparent truth while maintaining its falsity all come up short. I conclude that appearances are not deceiving, and (2) actually is true. Since (3) is also true, causal reductionism and strong causal eliminativism are false. Additionally, we saw above that weak causal eliminativism is only as plausible as the Davidson-Strevens thesis, since it is this thesis which allows the weak eliminativist to avoid the absurd conclusions that 'carbon emissions caused climate change' is unassertible. Since the Davidson-Strevens thesis falters, so too does weak causal eliminativism. This leaves causal emergentism as the last position standing.

Of course, additional epicycles are always available; additional semantic or pragmatic theses could be introduced to pick up the slack left over by the Lewis and Davidson-Strevens theses. (As I said above, it is a criterion of bare adequacy on such an account that it provide some explanation of why it is that we would choose to talk in accord with such semantic or pragmatic theses.) It cannot be predicted in advance whether counterexamples to these

further theses could be discovered. More effective than arguing against these epicycles, I think, is demonstrating that there is a more attractive alternative. And, as the example of Lewis's 1973 account demonstrates, in order to be causal emergentists, we need not deny that high-level causal facts reduce to, or obtain in virtue of, fundamental physical facts. We need only deny that they obtain exclusively in virtue of low-level *causal* facts. Such a metaphysical commitment is far from "extravagant", as Strevens claims. And theories of causation which entail causal emergentism are already well established and well regarded. Besides the account of Lewis (1973), by the way, causally emergentist theories of causation include the process theory of Dowe,[42] the regularity theory of Mackie,[43] the probabilistic theory of Suppes,[44] and the interventionist theory of Woodward.[45] (Or, at least, these theories generate causal emergentism is we are allowed to feed both high- and low-level events into their truth-conditions.) Causal emergentism allows us to account for the apparent truth of both (2) and (3) without the contortions of elaborate semantic or pragmatic theses. And it can do so without rustling any (microphysical) reductionist feathers. To my mind, this counts as a considerable point in its favor.

References

Collins, John, Ned Hall & L. A. Paul, editors. 2004. *Causation and Counterfactuals*. The MIT Press, Cambridge, ma. [28]

Davidson, Donald. 1967. "Causal Relations." *The Journal of Philosophy*, vol. 64 (21): 691–703. Reprinted in *Essays on Actions and Events* (2001), Oxford University Press, 2nd edition, pp. 149–162. Page numbers are from *Essays on Actions and Events*. [8]

[42] Though I do not have the space to go into it here, the curious reader will find that Dowe more or less explicitly embraces causal emergentism in chapter 7 of Dowe (2000).

[43] Simply because c is part of a minimally sufficient condition for e, this does not entail that C is a part of a minimally sufficient condition for E, as the case of Gliese 163c and stagflation readily illustrates.

[44] This is because it is possible for both $\Pr(e \mid c) \neq \Pr(e \mid \neg c)$ and $\Pr(E \mid C) = \Pr(E \mid \neg C)$ to be true, if we understand C and E to be events in the measurable space over which Pr is defined, and c and e to be events such that $c \subsetneq C$ and $e \subsetneq E$. (For the details of Suppes' theory, see Suppes (1970)).

[45] Simply because there is an in-principle intervention on the maximally-fine-grained variable describing the fundamental physical state of Gliese 163c which brings about a change in the value of the maximally-fine-grained variable describing the fundamental physical state of stagflation, this does not mean that there is an in-principle intervention on a more coarse-grained variable describing Gliese 163c's orbit which brings about a change in the value of a coarse-grained variable describing whether or not the U.S. economy stagflates during the 1970's.

Dowe, Phil. 2000. *Physical Causation*. Cambridge University Press, Cambridge. [20], [27]

Field, Hartry. 2003. "Causation in a Physical World." In *Oxford Handbook of Metaphysics*, Michael Loux & Dean Zimmerman, editors, 435–60. Oxford University Press, Oxford. [17]

Kim, Jaegwon. 1976. "Events as Property Exemplifications." In *Action Theory*, Myles Brand & Douglas Walton, editors, 159–77. D. Reidel, Dordrecht. Republished in Varzi & Casati (1996, p. 117–136). [5]

—. 1984a. "Epiphenomenal and Supervenient Causation." *Midwest Studies in Philosophy*, vol. 9 (1): 257–270. [5]

—. 1984b. "Supervenience and Supervenient Causation." *The Southern Journal of Philosophy*, vol. 22 (S1): 45–56. [5], [16]

Lewis, David K. 1973a. "Causation." *The Journal of Philosophy*, vol. 70 (17): 556–567. [2]

—. 1973b. *Counterfactuals*. Blackwell Publishers, Malden, MA. [2]

—. 1979. "Counterfactual Dependence and Time's Arrow." *Noûs*, vol. 13 (4): 455–476. [2], [3], [10]

—. 1983. "New Work for a Theory of Universals." *Australasian Journal of Philosophy*, vol. 61 (4): 343–377. [10]

—. 1986. "Events." In *Philosophical Papers*, vol. II, 241–269. Oxford University Press, New York. [2], [10], [11]

—. 1994. "Humean Supervenience Debugged." *Mind*, vol. 103 (412): 473–490. [10]

—. 2000. "Causation as Influence." *The Journal of Philosophy*, vol. 97 (4): 182–197. Reprinted in Collins et al. (2004, pp. 75–106). [3], [22]

—. 2004. "Causation as Influence." In Collins et al. (2004), chap. 3, 75–106. [22]

Loewer, Barry. 2001. "Review of *Mind in a Physical World*." *Journal of Philosophy*, vol. 98 (6): 315–324. [17]

Mackie, John L. 1965. "Causes and Conditions." *American Philosophical Quarterly*, vol. 2 (4): 245–55. [19]

Maudlin, Tim. 2007. "A Modest Proposal Concerning Laws, Counterfactuals, and Explanations." In *The Metaphysics within Physics*, 5–49. Oxford University Press, Oxford. [19]

McLaughlin, Brian P. 1992. "The Rise and Fall of British Emergentism." In *Emergence or Reduction? Essays on the Prospects of Nonreductive Physicalism*, Ansgar

Beckermann, Hans Flohr & Jaegwon Kim, editors, 49–93. Walter de Gruyter, Berlin. [9]

Menzies, Peter. 1988. "Against Causal Reductionism." *Mind*, vol. 97 (388): 551–574. [1]

Ney, Alyssa. 2009. "Physical Causation and Difference-Making." *The British Journal for the Philosophy of Science*, vol. 60: 737–764. [17], [21]

Price, Huw. 1992. "Agency and Causal Asymmetry." *Mind*, vol. 101 (403): 501–520. [6]

Price, Huw & Richard Corry, editors. 2007. *Causation, Physics, and the Constitution of Reality: Russell's Republic Revisited*. Clarendon Press, Oxford. [17]

Russell, Bertrand. 1912. "On the Notion of Cause." *Proceedings of the Aristotelians Society*, vol. 13: 1–26. [7], [17]

Salmon, Wesley. 1994. "Causality without Counterfactuals." *Philosophy of Science*, vol. 61 (2): 297–312. [20]

Schaffer, Jonathan. 2001. "Causation, Influence, and Effluence." *Analysis*, vol. 61 (1): 11–19. [24]

Strevens, Michael. 2003. *Bigger than Chaos: Understanding Complexity through Probability*. Harvard University Press, Cambridge, ma. [24]

—. 2008. *Depth: An Account of Scientific Explanation*. Harvard University Press, Cambridge, ma. [2], [8]

Suppes, Patrick. 1970. *A Probabilistic Theory of Causality*. North-Holland Publishing Company, Amsterdam. [27]

Varzi, Achille C. & Roberto Casati, editors. 1996. *Events*. Dartmouth, Aldershot. [28]

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press, Oxford. [21]