Lessons from the
Context Sensitivity of Causal Talk

Eric Swanson
ericsw@umich.edu

Suppose we have a theory of singular causation according to which

(1)    Caesar's birth was a cause of his death.

is true.[1]  Charge: It offends common sense to say that Caesar's birth was a cause of his death.  Response: The assertibility conditions of causal claims are affected by conversational context.  Even if (1) is true, in normal contexts it will be uninformative, or misleading, or not a suitable answer to the sorts of questions we are interested in.  And general pragmatic principles explain why it would offend common sense to assert even true sentences that are uninformative, misleading or not topical.  So it is no mark against a theory of causation that it predicts that (1) and certain other odd sounding sentences are true.

    This response, prominent in the work of Lewis (1973a, 558–559; 2000, 196), Bennett (1995, 130–133), and others, is based on the plausible idea that some distinctions made in natural language need not—indeed, should not—be reflected in metaphysics.  Natural language does distinguish between Caesar's birth and Brutus's stabbing, with respect to being a cause of Caesar's death, but perhaps our metaphysics of causation should not.  If we pursue this line, as I think we should, then we must ask which natural language distinctions do constrain our metaphysics, and how.  These questions are especially important for distinctions that are sensitive to features of conversational context, because we should not inadvertently impute the effects of such context sensitivity to our metaphysics.

    This paper starts by arguing that ordinary causal talk is far more sensitive to conversational context than has been recognized to date.  I then formulate a principle that helps characterize that context sensitivity.  I argue that this principle explains why some putative overgenerated causes are never felicitously counted, in conversation, as causes, and I argue that a plausibly strengthened version of the principle explains at least some of the oddness of 'systematic causal overdetermination.'  These explanations are a natural extension of the line that Lewis, Bennett, and others take with (1): when we are confronted with linguistic data that threaten to make trouble for our metaphysics, we try to give a plausible explanation of the data that does not require any changes to our metaphysics.  When we are successful, we are not obligated to change the metaphysics.  The linguistic explanations of (1) and many other examples are often seen as a boon to metaphysics—a

    [1]MACKIE 1980, LEWIS 1973a and 2000, and BENNETT 1988 offer such theories.

ground-clearing prerequisite to serious theorizing about causation. But the kind of context sensitivity I consider here has a worrying flip side. The metaphysics of causation turns out to be much less directly constrained by ordinary language judgments than we might have expected.

## 1.  The context sensitivity of causal talk

It is often noted that which events count as the causes of another event, in a conversation, is a dramatically context sensitive matter. In light of many philosophers suggest that our intuitions about causation should not be influenced by judgments about sentences of the form '$c$ was among the causes of $e$.' We should instead restrict our attention to putatively less context sensitive sentences of the form '$c$ was a cause of $e$.' This suggestion is underwritten by a tacit argument by analogy. Which books count as the books, in a conversation, is a dramatically context sensitive matter. But even if what counts as a book is a little context sensitive, 'a book' is much less context sensitive than 'the books.' By analogy, 'a cause of $e$' is much less context sensitive than 'the causes of $e$.' Indeed, it is sufficiently insensitive to context to be a suitable guide when we work on the metaphysics of causation. This argument is seductive, but it is specious, at least because the analogy between 'a book' and 'a cause of $e$' fails.[2]

The analogy fails because a given definite description of the form 'the $F$s of DEFINITE NP' is generally much more specific than its 'the $F$s' counterpart: 'the friends of Liem,' for example, is much more specific than 'the friends.' This specificity tempers the impact that facts about contextual salience can have on the interpretation of 'the $F$s of DEFINITE NP.' So definite descriptions whose nominals have genitive modifiers are generally less context sensitive than their counterparts without such modifiers—unless there is another reason for their context sensitivity.[3] (Continuing the example, 'the friends of Liem' is less context sensitive than 'the friends.') This suggests that the dramatic context sensitivity of 'the causes of $e$' may not be fully accounted for by the fact that it is a definite description. And if its context sensitivity has another source, then it would not be surprising to find that 'a cause of $e$' is context sensitive, too.

In fact this is what we do find. Unfortunately, moreover, neglect of the context sensitivity of 'a cause of $e$' has led our theorizing about causation astray. To bring this out I

---

[2]Lewis does concede that "even 'a cause of' may carry some hint of selectivity" (1986, 216). In his 1977 and 1984, Unger argues that the verb 'cause' and "other transitive causal verbs" are context sensitive, but does not discuss 'a cause of $e$.' See PEARL 2000, 222–223 for contrast.

[3]It is interesting that most of Russell's examples of definite descriptions have nominals with genitive modifiers. He singles them out as "descriptive functions," including "the father of $x$," "the sine of $x$," "the present King of France," "the author of *Waverley*," "the centre of mass of the Solar System at the first instant of the twentieth century" and so on (1905, 35; 1919, 323). Definite descriptions of this form seem to encourage 'attributive' over 'referential' readings, in something like Donnellan's senses (1966).

want to look at some linguistic data that clearly should not be accounted for in our meta-physics. For this reason these data put nonnegotiable demands on the linguistic theory that interfaces between our metaphysics and our ordinary language judgments: the linguistic theory has to account for these data on its own. But as we will see, it is plausible that a theory powerful enough to do this work can also do work usually taken to be the metaphysician's responsibility.

For all I will say here, the kind of context sensitivity that matters for present purposes may or may not make a difference to the truth conditions of causal claims. What I am interested in for present purposes is how the assertibility conditions of sentences of the form '*c* was a cause of *e*' are sensitive to conversational context. By the *assertibility conditions* of a sentence I mean the circumstances in which it is appropriate for a speaker who knows all the relevant non-semantic facts to use that sentence. There is no doubt that the assertibility conditions of any given causal claim vary with conversational context, in the familiar way that the assertibility conditions of any sentence vary with conversational context: clearly it is often inappropriate to say rude things, or things that have already been said, or things that are manifestly obvious, and so on. What I want to call attention to here is one unnoticed way in which the assertibility conditions of '*c* was a cause of *e*' vary with conversational context.

Suppose I know that the leak caused the puddle, and that the puddle together with the cold caused the ice. I tell you about the ice, and you ask about its causes. I could begin to answer your question with any of:

(2)  The leak was a cause of the ice. (… So if we fixed the leak…)

(3)  The puddle was a cause of the ice. (… So if we fixed the drain…)

(4)  The cold was a cause of the ice. (… So if we fixed the heater…)

(5)  The leak was a cause of the ice, and the cold was a cause of the ice.

(6)  The puddle was a cause of the ice, and the cold was a cause of the ice.

But I could not appropriately say

(7)  #The leak was a cause of the ice, and the puddle was a cause of the ice.

(8)  #The puddle was a cause of the ice, and the leak was a cause of the ice.

Assertions of (7) and (8) would not appropriately describe the stipulated causal relations. Indeed, they would misdescribe them, conveying that while the leak and the puddle were both causes of the ice, the leak was not a cause of the puddle.

What is wrong with using (7) and (8) to describe the stipulated situation? Roughly speaking, (7) and (8) wrongly double-count a causal route to the ice. Two bits of termi-

nology will make it easier to give a more precise and more general explanation than this one. First, many philosophers of causation, including Lewis and Bennett, explicitly aim to characterize a "broad and non-discriminatory" causal relation (Lewis 1973a, 559). Such accounts say, for example, that throwing a switch so that a train continues down the right-hand track (instead of the left-hand track) bears this non-discriminatory causal relation to the train's arrival, even if the tracks rejoin before the arrival and the switching makes no difference to the time and manner of the arrival. There is significant disagreement about the nature of this relation, but I will assume that there is a metaphysical natural kind that Lewis, Bennett, and similar philosophers of causation are aiming to characterize. For convenience I will call this putative natural kind *causal relevance*. I leave open the question which analysis of causal relevance is right, and I also leave open the question which causal relata that are causally relevant to *e* can count as a cause of *e* in a given context. (Note that causal relevance is a directed relation, because the switch is causally relevant to the arrival but not vice versa.) Second, a (possibly infinite) sequence of causal relata $\langle \ldots, e_{n-2}, e_{n-1}, e_n \rangle$ constitutes a *causal path* to $e_n$ just in case $e_{n-1}$ is causally relevant to $e_n$, $e_{n-2}$ is causally relevant to $e_{n-1}$, and so on. It will be important to confine our attention to causal paths that are *maximal* in the sense that they have no proper super-sequences with the same terminal event. The "broad and non-discriminatory" nature of causal relevance means that there will be very many causal paths to any effect, and that very many of those paths will overlap. It is because of this overlapping that I use the term 'causal path' instead of the more familiar 'causal chain': on my definition distinct causal paths to *e* can overlap by sharing events other than *e*. Some intuitive understandings of the 'chain' metaphor obscure this possibility.

I will appeal to the following principle in explaining why (7) and (8) misdescribe the situation we have been considering:

USE GOOD REPRESENTATIVES

> When you ascribe some causal responsibility for *e* to a causal path to *e*,
> use good representatives of that path for the purposes at hand.[4]

The refinement of this principle would be an illuminating project, but fully characterizing the kind of context sensitivity it approximates would require intensive empirical investigation. My intention here is to sharpen the principle just enough to get at some ways in which our theorizing about causation should be informed by the context sensitivity of causal talk.

---

[4] A serious discussion of absences and omissions would take us very far afield, but I want to note that it is easy to generate analogues of the leak/puddle case for absences. (This is a 'new' kind of context sensitivity, not obviously related to the broadly normative context sensitivity we are accustomed to seeing in causal talk about absences.) USE GOOD REPRESENTATIVES might explain these cases as well: for example, 'absence' expressions might represent merely possible causal paths.

Before I explain how USE GOOD REPRESENTATIVES sheds light on the infelicity of (7) and (8), I need to say a bit about the thought that we ascribe causal responsibility not to particular events but to causal paths, and a bit about the sense in which some but not all events on a causal path are "good representatives" of that path. First, in saying '$c$ was the cause of $e$' or '$c$ was causally responsible for $e$' we almost never mean that $c$ alone was causally responsible for $e$. We know, after all, that the other events on the causal paths through $c$ to $e$ are partly responsible for bringing about $e$, in the sense that $e$ would have been prevented, or would have happened in a different way, if any of many of the events on the causal paths through $c$ to $e$ had been prevented or had occurred differently than they did. But for many purposes little or nothing would be gained by mentioning multiple events on a causal path to $e$. Often surprisingly few 'landmarks' suffice to enable us to pick out the features of represented causal paths that are relevant to a particular conversation. So it is often most efficient to pick representatives of causal paths in conformity with the following constraint: they should provide conversational participants with enough information about the represented causal paths for the purposes at hand. And (as ever) speakers should mind the costs of adding misleading or unnecessary information to a conversation. Ceteris paribus, the better an event does at satisfying this constraint, the better a representative it is. Of course, when predictable inferences would lead the addressee to importantly false conclusions, or when important features of the causal paths are obscure or surprising, the speaker may use more representatives (or other devices) to ensure that her addressee draws out the relevant features of a path with appropriate, accurate detail. But very often a speaker can achieve her conversational goals by picking out a single apt representative for some causal paths—often, an event that has moral significance, or could have been easily prevented, or is such that a change in it, holding certain aspects of the situation fixed, would have made differences relevant to the conversational participants.

Even given all this, one might worry that inevitably too many or too few things will count as "good representatives." More concretely, the fact that (2) and (3) are felicitous in many contexts shows that both the leak and the puddle can be good representatives of the causal paths through them to the ice. Because USE GOOD REPRESENTATIVES does not demand that we pick exactly one event to represent a causal path or paths, it might look powerless to explain the infelicity of (7) and (8). But this is not so. As our judgments about (2) and (3) show, we start in a context in which both the leak and the puddle are good representatives of the causal paths through the leak to the ice. For precisely this reason, we do not generally need to use both the leak and the puddle to represent paths through the leak to the ice well enough to satisfy USE GOOD REPRESENTATIVES. Once the leak is used to represent those paths, if it does so well enough to satisfy USE GOOD REPRESENTATIVES, the puddle becomes a poor representative of those same causal paths. To cite the puddle, too, as their representative would be to add information to the conversation

that has been rendered unnecessary enough to be misleading. Using the leak to represent the relevant causal paths makes otiose—mid-utterance—the representative role that the puddle otherwise could have played.

Use good representatives also explains why it is easy to hear (7) and (8) as describing a situation in which neither the leak nor the puddle is causally relevant to the other. This is because when a speaker uses both the leak and the puddle as representatives of some causal paths, she conveys that both the leak and the puddle are needed to successfully represent some causal paths—that neither plays a representative role successfully discharged by the other alone. From this point it is natural for the addressee to hypothesize that the leak and the puddle play distinct representative roles by representing distinct causal paths—that because the puddle is needed to represent causal paths that are not well represented by the leak, the puddle represents paths that do not even include the leak.

Of course, the selection of a particular representative for a causal path or paths is not irrevocable. But it is difficult to switch from one representative to another. This is similar to the difficulty in switching between particular restrictions on domains of quantification. Speakers regularly use 'everyone' not to talk about absolutely everyone but rather about every employee, or every child, or every addressee, or what have you. But once a particular restriction on 'everyone' is in place, considerable conversational pressure holds it there. The reason for this is that it takes group effort for conversational participants to coordinate on domains of quantification, and one should not ask conversational participants to engage in such an effort without good reason. Similarly, it takes group effort to coordinate on a new representative for a given causal path. The bar to switching between representatives means that to pick one causal representative is in effect to screen off other potential representatives to some degree, making it important to pick representatives of causal paths that will well serve both present and (perhaps unconceived) future conversational purposes. For all a participant in a typical conversation knows, future conversational purposes could well demand especially good representatives. So there is pressure to select representatives for a path that are not significantly bettered by other potential representatives—a pressure counterbalanced, as always, by the demands of conversational and cognitive economy.

One prima facie problem for my use of use good representatives is that it is easy to imagine sentences like (9)–(11) being felicitous descriptions of the situation we have been considering:

(9) The leak was a cause of the ice by being a cause of the puddle.

(10) The leak was a cause of the puddle, and because of that was a cause of the ice.

(11) The leak was a cause of the puddle, and the puddle was a cause of the ice, so the leak was a cause of the ice.

For example, we might use these sentences to explain why, given that we can fix neither the drain nor the heater, we ought to fix the roof. These sentences explicitly allot two representatives—the leak and the puddle—to the causal paths through the leak through the puddle to the ice. But this does not undermine my explanation of the infelicity of (7) and (8). Why might we allot two representatives to these paths? Consider some contexts in which we might actually use (9)–(11). Often, we use such sentences to explain why the leak was causally relevant to the ice, in speaking with someone who grants that the puddle was causally relevant but does not believe that the leak was, too. The causal paths that we are trying to inform our addressee about need (for these purposes) to be represented both by the leak and the puddle, for—as far as the addressee knows—the paths through the puddle to the ice do not include the leak. We inform addressees that some of those paths do include the leak by being explicit about the fact that we are using both the leak and the puddle to represent those paths.

Here is another situation that illustrates how USE GOOD REPRESENTATIVES can do helpful explanatory work. Suppose that the increasing heat caused both the expansion of the gas and the weakening of the container, and that the expansion and the weakening in turn jointly caused the rupture of the container. A fully informed speaker may appropriately describe this situation using any of (12)–(15).

(12)    The expansion of the gas was a cause of the rupture of the container.

(13)    The weakening of the container was a cause of the rupture.

(14)    The increasing heat was a cause of the rupture of the container.

(15)    The expansion of the gas was a cause of the rupture of the container. The weakening of the container was also a cause of the rupture.

But for many purposes (16) and (17) are sub-par (indeed, misleading) descriptions of the case.

(16)    The expansion of the gas was a cause of the rupture of the container. The increasing heat was also a cause of the rupture.

(17)    The weakening of the container was a cause of its rupture. The increasing heat was also a cause of the rupture.

For some purposes the increasing heat is a good representative of the causal paths through it to the rupture of the container. For other purposes, the expansion of the gas and the weakening of the container are good representatives of those paths. In this quite symmetric case, our purposes are often of this general type—they are purposes for which either the relatively 'distal' increasing heat or the relatively 'proximal' expansion of the gas and weakening of the container are crucial. But if the weakening was overdetermined—if it

would have occurred whether or not the heat had increased—(17) is likelier to sound felicitous. This is because the weakening would not counterfactually depend on the increasing heat, making the heat a somewhat misleading (and thus somewhat flawed) potential representative of the causal paths through the weakening to the rupture.

As I mentioned earlier, my goal here is not to fully characterize the notion of goodness at play in USE GOOD REPRESENTATIVES. I want to use USE GOOD REPRESENTATIVES only to help structure our explanations of how conversational context influences the assertibility conditions of causal claims. Why does this structure matter? Consider two questions:

1. Which events that are causally relevant to *e* can felicitously count as a cause of *e*, in which contexts?

2. Given a particular causal path to *e*, which events can felicitously represent that path, in which contexts?

Both these questions are very difficult, and both have (to say the least) a significant empirical component. We are nowhere near to having systematic answers to either of them. But the second question is less difficult than the first. Ceteris paribus it is much easier to compare how well two events that share a causal path to *e* do at representing that path, in a context, than it is to compare how well events on different causal paths do at counting as causes of *e*. As we will see, in certain important cases it is easy to see how various events fare in the competition to represent a given causal path.

## 2. Transitivity and overgeneration

A common move in the literature on causation, events, and causal transitivity is to argue that a theory overgenerates causes: it offends common sense by counting some event (or fact, or whatever) as a cause of another, when, intuitively, it is not. And so the theory is deemed false, or at least 'costly.'

In a sense we ought to worry about this kind of objection only to the extent that we aim to say what it is for one event to count as a cause of another in a particular conversational context. If our aim is simply to say what it is for one event to count as causally relevant to another, then the fact that some events that are causally relevant to *e* rarely if ever count as a cause of *e* may be utterly unsurprising. Consider Lewis's suggestion that his analysis of causation is really an analysis of "causal histories," parts of which "will not be at all salient in any likely context … : the availability of petrol, the birth of the driver's paternal grandmother, the building of the fatal road, the position and velocity of the car a split second before the impact" (1986, 215–216). If, with Lewis, we hypothesize that causation is transitive, then we can expand this list indefinitely. One part of the causal history of the building of the fatal road, for example, is the crew's painting of the north

curb line. So the crew's painting of the north curb line is part of the causal history of the accident. Granted, it rarely if ever counts as a cause of the accident. But (the arch proto-contextualist might say) so what? (Lewis makes similar maneuvers in his 1979 and 1996.) We as metaphysicians are interested in the metaphysics of the causal relevance relation—a metaphysics that would yield sufficient conditions for the assertibility of sentences of the form '*c* was a cause of *e*' only if it were supplemented with a sufficiently comprehensive semantics and pragmatics.

Whether one thinks this gambit is principled will depend, I think, on what one thinks of notions like causal relevance and Lewis's "causal history." To the extent that one thinks the philosophically interesting questions about causation are about a "broad and non-discriminatory" causal relation (Lewis 1973a, 559) one may welcome the move to '*c* was causally relevant to *e*.' Philosophers who suspect that we have little helpful pretheoretic grip on a broad, inclusive causal relation may be assuaged by seeing detailed explanations of why a particular event is a relatively bad representative of causal paths to *e*, even supposing that it is causally relevant to *e*.

Consider

SWITCH

> A train departs Mountain Station. It comes to a fork in the track, where an engineer flips a switch so that the train continues on the right-hand track. The right-hand and left-hand tracks rejoin before reaching Valley Station. The train arrives at Valley Station, and the switching made no difference to the time and manner of its arrival.

Ned Hall asks: "Is [the engineer's] flipping the switch a cause of the train's arrival? Yes, it is, though the opposing reaction surely tempts" (2000, 205). Hall gives a number of reasons to think the flip of the switch is a cause of the train's arrival—reasons that perhaps should be construed as showing only that the switch was causally relevant to the arrival. But we also want to know why the opposing reaction is so tempting.

USE GOOD REPRESENTATIVES helps us answer this question. By saying

(18)  #The engineer's flipping the switch was a cause of train's arrival.

a speaker would, by USE GOOD REPRESENTATIVES, convey that the flipping of the switch is a good representative of the causal paths running through it to the train's arrival. But using the switch to represent those paths gives a misleading picture of which events on the paths matter most for normal purposes. It is at least natural to infer, from the claim that *c* was a cause of *e*, that whether *c* occurred made some difference to the likelihood of *e* occurring, despite the fact that the counterfactual dependence of *e* on *c* is not necessary for *c*'s being causally relevant to *e*. It is plausible that this is because we are often more interested in counterfactual dependence than in causal dependence. We want to know what

would have made a difference; we want to know what consequences changes upstream would have had. (To be sure, causal dependence is often more prominent when ethical issues are relevant.) Events on the causal paths to the train's arrival that the arrival does counterfactually depend on do better, in this quite important respect at least, at representing those causal paths. And the flipping of the switch does not better the train's departure (say) in the other respects that might be relevant to the goodness of a representative. Thus the departure, among other events, is for many purposes a significantly better representative of the relevant paths than the flip is. So is the train's motion after the tracks rejoin. And if the engineer's flipping the switch were used to represent those paths, it would fill the representative role that the train's departure would better fill, thereby linguistically screening off the train's departure and other events that are better representatives of the relevant paths.

The problem is not that there is something wrong with screening off per se—it happens all the time. The problem, rather, is that many of the screened off candidates do better at representing the causal paths to the arrival than the flip. And as I argued earlier, it is bad form to choose a representative for a causal path if that representative would serve our evolving conversational purposes significantly worse than other candidates would. All this is consistent with the supposition that the flip is causally relevant to the arrival. It also does not mean that the flip is a bad representative of causal paths through it that end in events that occur while the train is on the right-hand track. Indeed, it is plausible that the flip will be a good representative of those causal paths, since those events do counterfactually depend on the flip.

This sort of explanation does not depend on there being a total preordering of the events that are causally relevant to $e$, in terms of their fitness for being counted as a cause of $e$. Indeed, I doubt that events can be ordered in this way. Is the spark a better or worse candidate to count as a cause of the fire than the presence of oxygen? Is my dropping the glass a better or worse candidate to count as a cause of its shattering than its fragility? Even relative to particular contexts such questions do not always have good answers. So there are many pairs of events that are both causally relevant to some event $e$ but that are not comparable with respect to their fitness for being counted as among the causes of $e$. What the explanation does rely on is the claim that given a particular causal path some events are better representatives of it than others. When we ask, of the causal paths that run through the flipping of the switch to the arrival of the train, whether some events better represent them than the flipping of the switch, the answer is 'yes': after all, there is the departure, the train's movement after the tracks rejoin, the entire course of the train before it arrives, and so on.

What goes into making an event a good representative of a causal path is an important and difficult question. And some philosophers of causation have arguably already made significant progress on it, under a different mode of presentation. When they proffer

analyses that do better than "broad and non-discriminatory" analyses at matching our intuitions about what can count as a cause of what, they have taken themselves to be doing metaphysics. But a philosopher with a non-discriminatory theory of causation can reconstrue such work as providing resources that help her say which events count as good representatives of a causal path. Because conversational context helps determine which events are good representatives of a causal path, every metaphysical theory of causation needs a 'good representative' theory to mediate between it and our ordinary language judgments. The question is just how much work that theory should do, and how much work the metaphysics should do. Even if a 'broad' metaphysics of causation needs a more ambitious theory of good representatives than a 'narrow' metaphysics, the need for this kind of supplementation does not obviously cut against the broad metaphysics. So the advocate of a broad metaphysics can simply integrate techniques of a narrower putative 'metaphysics' of causation to work in her linguistic theory of good representatives.

I do not want to pursue this line further here, however, because I think it is illuminating to see how much work USE GOOD REPRESENTATIVES can do without being tailored to a particular metaphysics. Hartry Field's BOMB case, like SWITCH, nicely displays its potential.

BOMB

> Billy plants a bomb in a room. Suzy comes into the room, notices the bomb, and flees. Suzy later has a checkup and is found to be in perfect health.

The presence of the bomb is a cause of Suzy's fleeing, and Suzy's fleeing is a cause of her perfect health the next day. But "the bomb is not a cause" of Suzy's health, making this a "counterexample to transitivity" (YABLO 2004, 119). Or so the story goes.

It is very important to be clear about what needs to be explained. The presence of the bomb can in fact be cited as a cause of Suzy's health, as long as it is also counted as a cause of her fleeing the room, or of her not being present at the time of the explosion. For example: the presence of the bomb at time $t$ caused Suzy to believe that an explosion was imminent. And given that an explosion occurred soon after $t$, Suzy should be glad that the bomb was present. Indeed, the presence of the bomb was a cause of her good health, because the presence of the bomb was a cause of her fleeing before the explosion. Of course we are sketching causal paths to Suzy's health using multiple events, as is often necessary when the causal relevance of one such event is in question. We saw this earlier with (11):

(11)   The leak was a cause of the puddle, and the puddle was a cause of the ice, so the leak was a cause of the ice.

So (19) is not particularly odd.

(19)   The presence of the bomb was a cause of Suzy's fleeing, and (given that an

explosion occurred) Suzy's fleeing was a cause of her good health, so the presence of the bomb was a cause of Suzy's good health.

Our real task, then, is to explain the oddness of

(20)   #The presence of the bomb was a cause of Suzy's good health.

uttered with no further elaboration on the details of the case.

In principle, the presence of the bomb might represent causal paths through the explosion, and it might represent paths through Suzy's fleeing. Paths of the first kind are irrelevant because Suzy's fleeing makes the explosion causally irrelevant to her health. (To be sure, the explosion would have been causally relevant to Suzy's injury, but for the fortunate fact that she fled the room and prevented herself from being injured.) And the presence of the bomb does not represent paths through Suzy's fleeing to her health very well compared to other events on those paths. For example, her fleeing the room does better at representing the causal paths through her fleeing the room to her subsequent health than the presence of the bomb would, because (again) counterfactual dependence matters so much to us. If Suzy had stayed in the room she would not have been healthy the next day; if the bomb had not been present there would have been no explosion, so whether or not the bomb had been present Suzy would have been healthy the next day. By uttering (20) on its own a speaker screens off potential representatives of some causal paths to Suzy's perfect health, among them her fleeing the room—a significantly better representative of the relevant causal paths. So in most contexts, her fleeing the room can count as a cause of her health, but the presence of the bomb cannot. Whether the presence of the bomb is causally relevant to Suzy's good health is of course not dependent on context, but whether the presence of the bomb counts as a cause of Suzy's good health does depend on context. Generally it does not, but when the speaker does not let the presence of the bomb screen off her fleeing the room (as in (19)) it may.

Bomb is an especially interesting case for present purposes because it elicits diverging intuitions about causal relevance. As a matter of fact, Hall thinks that the bomb is not causally relevant to Suzy's health (p.c.); Lewis thinks it is (2000, 194).[5] It is not surprising, I think, that there are many disagreements about what is causally relevant to what. The chief difficulty here is that natural language will not always help resolve such disagreements.

---

[5]Lewis has at least two rationales here. First, the bomb's presence causes Suzy's health to be caused in one way rather than another, so the bomb's presence causally influences the causal history of Suzy's health. As a result it is part of the causal history of Suzy's health (Lewis 2000, 194–195). Second, if "whole" causal explanations are "the biggest chunk of explanatory information that is free from error" (1986, 218), then Suzy's fleeing is part of the causal explanation of her health. It would be obscure why her fleeing is explanatorily relevant if the bomb's presence was not also part of the causal explanation of her health. So the bomb's presence must be part of the explanation too. But if causal histories just are "whole" causal explanations (as Lewis sometimes suggests (1986, 218–219)), then the causal history of Suzy's health must include the bomb's presence as well.

Our judgments about sentences like (20) do not constrain the metaphysics of causation as directly as many have thought.

Lewis's line on BOMB is not vindicated, of course, by the fact that he can explain why we do not usually count the bomb as a cause of Suzy's health. But the existence of such an explanation supplies a principled way for him to agree with common sense that the bomb counts as a cause of Suzy's health in relatively few contexts, while still maintaining that the bomb is causally relevant to Suzy's health. So in response to a philosopher who claims that BOMB is a counterexample to causal transitivity simpliciter (YABLO 2004, 119), or that such cases are counterexamples to the conjunction of transitivity and the sufficiency of counterfactual dependence (HALL 2004, 246–248), Lewis could observe that the planting of the bomb is not, in ordinary contexts, a good representative of the causal paths through it to Suzy's health. USE GOOD REPRESENTATIVES affords a metaphysically neutral way for any theorist of causation to explain away our linguistic judgments about cases like BOMB.

## 3. Overdetermination

Sometimes two otherwise innocuous looking causal claims seem odd when taken together. Earlier we considered circumstances and contexts in which (2) and (3) are both appropriate on their own, but (7) is odd.

(2)   The leak was a cause of the ice. (… So if we fixed the leak…)

(3)   The puddle was a cause of the ice. (… So if we fixed the drain…)

(7)   #The leak was a cause of the ice, and the puddle was a cause of the ice.

I explained this oddness in terms of a context change in the midst of (7). The first conjunct of (7) changes the conversational context by making the puddle an otiose (and thus poor) representative of the paths through it to the ice. Evaluated relative to that context, the second conjunct of (7) is most naturally interpreted as representing causal paths to the ice that do not include the leak. (7) is inappropriate at least to the extent that it is inappropriate to suggest that the leak and puddle are both needed to represent different causal paths to the ice, for speakers make that suggestion when they assert (7).

Context shifts within conjunctions are not at all unusual. In particular, in many contexts sentences that could be felicitously asserted in isolation would be infelicitous if conjoined or asserted together. Suppose for example that yesterday in the park I saw two dogs—one the largest I have seen in months, and the other the smallest I have seen in months. I tell you that I saw a dog in the park yesterday. I could then felicitously say either (21) or (22).

(21)   The dog was the largest I've seen in months.

(22)    The dog was the smallest I've seen in months.

But (unless it has been months since I have seen any other dogs) I could not felicitously say

(23)    #The dog was the largest I've seen in months, and the dog was the smallest I've seen in months.

The first conjunct of (23) changes the context in a way that makes it inappropriate to say the second conjunct: it makes the larger dog significantly more salient than the smaller dog, thereby affecting the interpretation of second occurrence of 'the dog.' This is obviously not in conflict with the fact that (22) can be used, on its own, to make the smaller dog more salient than the larger dog. So by appealing to context change we can hold onto our intuitive judgments about (21), (22), and (23) without questioning the soundness of (appropriately restricted) conjunction introduction.

Skeptical of the force of arguments from the threat of systematic overdetermination, Ted Sider writes:

> Should we say that *a baseball* caused a certain window to shatter? Or the *the parts of the ball* caused the window to shatter? Or that *the event of the ball's striking the window* caused the window to shatter? Or that *the fact that the ball struck the window* caused the window to shatter? Or something else? One wants to say *all* of these things! That is certainly the natural view. (2003, 720)

I agree with Sider that the natural view is that all of these things are fine to say on their own. Nevertheless their conjunction is undeniably odd:

(24)    #The baseball caused the shattering, its parts caused the shattering, the event of its striking the window caused the shattering, and the fact that the ball struck the window caused the shattering.

(24) sounds odd for a by now familiar reason: it wrongly suggests that the baseball, its parts, and all the rest are on different causal paths to the shattering of the window. It wrongly suggests that, like soldiers on a firing squad, the baseball, its parts and so on are causally separate from each other in the sense that there are differences between what is causally relevant to them, between what they are causally relevant to, or both.

One might find overdetermination of the sort evoked by (24) to make for such "an ugly picture" (MERRICKS 2001, 67) that we are obligated to reject at least some of its conjuncts. Merricks, for example, avoids systematic overdetermination in part by holding that "If the baseball exists, it does not cause the shattering of the window" (57). But because it is plausible to explain the oddness of (7) by appealing to a context change in its

midst—rather than by rejecting (2) or (3)—we are amply justified in asking whether the oddness of (24) and sentences like it could be due to similar context changes. Unless he can show that there is no such context change in (24), Merricks simply is not licensed to conclude that we should reject any of its conjuncts.

On my own positive view about this case, there are numerically distinct sufficient causes of the shattering, in competition with each other only qua potential representatives of causal paths. Let a *causal cluster* be a set of causal relata $E$ such that exactly the same causal relata are causally relevant to all the members of $E$, and all the members of $E$ are causally relevant to exactly the same causal relata. We can define one causal cluster's being causally relevant to another in terms of causal relevance between their members (and we can rely on context to indicate which sense of 'causally relevant' is in play). One causal cluster is *causally relevant* to another just in case all the members of the first are causally relevant to all the members of the second. A (possibly infinite) sequence of causal clusters $\langle \ldots, E_{n-2}, E_{n-1}, E_n \rangle$ constitutes a *causal cluster path* to $E_n$ just in case $E_{n-1}$ is causally relevant to $E_n$, $E_{n-2}$ is causally relevant to $E_{n-1}$, and so on. As before, we confine our attention to causal cluster paths that are *maximal* in the sense that they have no proper supersequences with the same terminal causal cluster. Accordingly, we have

STRENGTHENED USE GOOD REPRESENTATIVES

> When you ascribe some causal responsibility for $E$ to a causal cluster path to $E$, use good representatives of that path for the purposes at hand.

We can account for the leak/puddle case with USE GOOD REPRESENTATIVES; STRENGTHENED USE GOOD REPRESENTATIVES is not necessary. But because the latter applies to singleton causal clusters, it can be used to explain any data explained by the former.

If wholes, parts, events, facts, objects, and so on all exist, then even granting that they are distinct from each other it is plausible that many of them are members of the same causal clusters and so compete to represent the same causal cluster paths. It is plausible, for example, that exactly what is causally relevant to the baseball is causally relevant to the atoms arranged baseballwise, and that exactly what the baseball is causally relevant to is what the atoms arranged baseballwise are causally relevant to. This is so even granting that the baseball is distinct from the atoms arranged baseballwise, for it simply does not follow from the distinctness of $x$ and $y$ that $x$ and $y$ are causally relevant to different things, or that different things are causally relevant to $x$ and $y$. So STRENGTHENED USE GOOD REPRESENTATIVES can also explain the oddness of (24). The form of the explanation is by now familiar: if we have already used the baseball's atoms to represent causal cluster paths to $E$, then we have made otiose the representative role that the baseball previously could have played.

Some philosophers think causal powers indicate what exists, in a way that might seem to make trouble for the explanation just offered. For example, Jaegwon Kim argues that

"To be real … is to have causal powers; *to be real, new, and irreducible, therefore, must be to have new, irreducible causal powers*" (1993, 204). This particular argument is invalid: to be uncommon, for example, is not necessarily to have uncommon causal powers.[6] Other methods might suffice to establish Kim's conclusion or something like it, however. Sydney Shoemaker also thinks that causal powers play an individuating role, contending that

> … [W]hat makes a property the property it is, what determines its identity, is its potential for contributing to the causal powers of the things that have it. This means, among other things, that if under all possible circumstances properties $X$ and $Y$ make the same contribution to the causal powers of the things that have them, $X$ and $Y$ are the same property. (1980, 212)

I doubt that views like these make any trouble for my explanation of (24)'s oddness. By distinguishing $x$'s causal susceptibilities from what is causally relevant to $x$, and distinguishing $x$'s causal powers from what $x$ is causally relevant to, we can say that $x$ and $y$ may have distinct causal susceptibilities and powers even if they are members of the same causal cluster. One causal susceptibility of alcohol, for example, is its tendency to burn when exposed to open flame in the presence of oxygen. One causal power of alcohol is its ability to dissolve shellac. A quantity of alcohol has this causal susceptibility and causal power whether or not it ever actually burns as a result of exposure to open flame or is ever actually causally relevant to any dissolvings of shellac. More generally, membership in the same causal cluster is entirely a matter of what is causally relevant to what in the actual world. Causal susceptibilities and causal powers are not.

The importance of these distinctions is especially clear when we consider some intuitive differences between mental events and correlated physical events. Suppose the pain and the C-fiber firing are part of the same causal cluster. There might nevertheless be a difference between what would have happened if the pain had occurred without the C-fiber firing and what would have happened if the C-fiber firing had occurred without the pain, and such a difference could well reflect differences between the causal powers of pain and the causal powers of C-fiber firing. There are many ways to flesh out the details here. Suppose for sake of argument that Lewis's 1973a account of causation, reconstrued as an account of causal relevance, is right. Since Stalnaker 1968 and Lewis 1973b, the most plausible semantics for counterfactuals have not validated antecedent strengthening. That is, those semantics allow for cases in which

1. If there were to be a C-fiber firing, the man would wince, and if there were not to be a C-fiber firing, the man would not wince. (So C-fiber firing, if it occurs, counts as a cause of the man's wincing on Lewis's 1973a account.)

---

[6]Moreover, one might think Kim should allow that some "real" things are distinct in virtue of differences in their causal susceptibilities but not their causal powers.

2. If there were to be a pain, the man would wince, and if there were not to be a pain, the man would not wince. (So pain, if it occurs, counts as a cause of the man's wincing on Lewis's 1973a account.)

3. If the neurological basis of pain had not been C-fiber firing and there were to be a C-fiber firing, the man would not wince, and if the neurological basis of pain had not been C-fiber firing and there were not to be a C-fiber firing, the man would not wince.

4. If the neurological basis of pain had not been C-fiber firing and there were to be a pain, the man would wince, and if the neurological basis of pain had not been C-fiber firing and there were not to be a pain, the man would not wince.

In many such cases, if a C-fiber firing and pain actually occur they will be members of the same causal cluster. Nevertheless they will not have the same causal powers: even if it were to have a very different neurological basis from what it actually has, a pain could still cause a wince. So there is ample room for a nonreductive physicalist to hold that a mental event is a member of some causal cluster that also contains a physical event, while holding that some mental events have causal powers distinct from the causal powers of any physical event.

## 4. Context sensitivity and the methodology of metaphysics

We started by looking for an illuminating story about 'causation itself'—a metaphysical natural kind the character of which we took to be independent of us, independent of the contingencies of our causal talk, and independent of our concept or concepts of causation. I argued that our ordinary causal claims are influenced by conversational context in significant ways, and so we turned to a not quite ordinary term—'causal relevance'—in the hope that it denotes the fundamental causal relation in a context free way. But then we found that in certain cases philosophers disagree about what even counts as causally relevant to what—let alone which is the true theory of causal relevance. And I argued, further, that in some cases judgments about ordinary causal talk do not indicate what counts as causally relevant to what, because we can explain those judgments in principled but metaphysically neutral ways. In effect, I showed that once we begin to flesh out the theory that interfaces between our metaphysics of causation and our causal talk, there is a strong possibility that our causal talk will grossly underdetermine the metaphysics of causal relevance.

We should not be surprised to arrive at this point. Given that we need complementary theories—one metaphysical, one linguistic—to get substantive predictions about our

causal talk, it would be fortuitous if our causal talk came anywhere close to uniquely determining the metaphysics of causation. In some respects this underdetermination is liberating, because we no longer have to worry about pesky cases like BOMB. But it also raises troubling methodological questions: if linguistic judgments dramatically underdetermine our metaphysics of causation, exactly what will help determine it?

For the time being I think we are justified in putting such skeptical worries to the side. Having a better sense of what work can be done by our linguistic theory provides us a better sense of where to look for examples that are genuinely probative for metaphysics. That said, whether a linguistic judgment is probative for metaphysics depends on the specific ways in which we explain the context sensitivity of causal talk. So we cannot make much real progress on the metaphysics of causal relevance without better understanding causal talk. In addition, some of the overlooked features of causal talk raise new questions for metaphysics. For example, I think we should investigate the possibility that causes are best thought of not as particular events (or facts, or whatever) but rather as causal paths themselves. (Note that taking causal paths to be the causal relata is not necessarily to hold a "process" view of causation: one needn't be committed to "conserved quantities" in the sense of DOWE 1992 and SALMON 1994.) These are roughly instrumental reasons to be interested in the context sensitivity of causal talk: understanding it may help us hone the knife with which we try to carve causal reality at its joints.

But I want to warn against the thought that the study of causal talk is just instrumentally important—that we would do well to ignore it if we could only find a way to theorize about causal relations without the intermediary of judgments about causal claims. For an analogy, consider your initial, unarticulated philosophical curiosity about the nature of friendship. The property of being a friendly acquaintance of $P$ is broad and non-discriminatory, instantiated at least by anyone who in some conversational context counts as a friend of $P$. We can know quite a lot about this relation without knowing anything about the ways in which 'is a friend of' is sensitive to conversational context. But saying what it is for two people to be friendly acquaintances obviously does little to address our curiosity about friendship. Ignoring context sensitivity makes our task easier—it is clearly easier to satisfy our philosophical curiosity about friendly-acquaintance-ship than it is to satisfy our philosophical curiosity about friendship—but ease of theorizing does not warrant such a change in subject. Any philosophically respectable course here will have to deal with or work around the context sensitivity of 'is a friend of.'

Similarly, we cannot ignore the context sensitivity of causal talk without neglecting much that is of philosophical interest. This is because our unarticulated curiosity about causation is in part a curiosity about causal thinking, which is crucial to folk psychology, moral judgment, scientific reasoning, and a host of other philosophically rich topics. It is impossible to cleanly excise ordinary causal talk from ordinary causal thinking, so to study one is to study the other. And we should welcome this connection: to take just one

example, progress on the theory of good representatives would likely help us better understand the connections between normative and causal judgments. It is rarely cheering to see that we cannot make progress on a family of philosophical problems without better understanding some related conversational context sensitivity. But we should not assume that we can satisfy our initial curiosity about causation without such an understanding.

## References

Bennett, Jonathan. 1988. *Events and Their Names*. Hackett Publishing Company, Indianapolis.

—. 1995. *The Act Itself*. Oxford University Press, Oxford.

Collins, John, Ned Hall & L. A. Paul, editors. 2004. *Causation and Counterfactuals*. MIT Press, Cambridge.

Donnellan, Keith. 1966. "Reference and Definite Descriptions." *Philosophical Review*, vol. 75: 281–304.

Dowe, Phil. 1992. "Wesley Salmon's Process Theory of Causality and the Conserved Quantity Theory." *Philosophy of Science*, vol. 59: 195–216.

Hall, Ned. 2000. "Causation and the Price of Transitivity." *Journal of Philosophy*, vol. 97 (4): 198–222. Reprinted in Collins et al. (2004), 181–203.

—. 2004. "Two Concepts of Causation." In Collins et al. (2004), 225–276.

Kim, Jaegwon. 1993. "The Nonreductivist's Troubles with Mental Causation." In *Mental Causation*, John Heil & Alfred Mele, editors, 189–210. Oxford University Press.

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

—. 1973b. *Counterfactuals*. Basil Blackwell Ltd., Malden, MA.

—. 1979. "Scorekeeping in a Language Game." In *Philosophical Papers*, vol. 1, 233–249. Oxford University Press, Oxford.

—. 1986. "Causal Explanation." In *Philosophical Papers*, vol. 2, 214–240. Oxford University Press, Oxford.

—. 1996. "Elusive Knowledge." In *Papers in Metaphysics and Epistemology*. Cambridge University Press, Cambridge.

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

Mackie, J. L. 1980. *The Cement of the Universe: A Study of Causation*. Oxford University Press, Oxford, second edn.

Merricks, Trenton. 2001. *Objects and Persons*. Oxford University Press, Oxford.

Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.

Russell, Bertrand. 1905. "On Denoting." In *Definite Descriptions: A Reader*, Gary Ostertag, editor, 35–49. MIT Press, Cambridge.

—. 1919. "Descriptions." In *Readings in the Philosophy of Language*, Peter Ludlow, editor, 323–334. MIT Press, Cambridge.

Salmon, Wesley C. 1994. "Causality without Counterfactuals." *Philosophy of Science*, vol. 61: 297–312.

Shoemaker, Sydney. 1980. "Causality and Properties." In *Identity, Cause, and Mind*, 206–233. Cambridge University Press, Cambridge.

Sider, Theodore. 2003. "What's So Bad About Overdetermination?" *Philosophy and Phenomenological Research*, vol. 67 (3): 719–726.

Stalnaker, Robert C. 1968. "A Theory of Conditionals." In *Ifs: Conditionals, Belief, Decision, Chance, and Time*, William L. Harper, Robert Stalnaker & Glenn Pearce, editors, 41–55. D. Reidel Publishing Company, Dordrecht.

Unger, Peter. 1977. "The Uniqueness in Causation." *American Philosophical Quarterly*, vol. 14: 177–188.

—. 1984. *Philosophical Relativity*. University of Minnesota Press, Minneapolis.

Yablo, Stephen. 2004. "Advertisement for a Sketch of an Outline of a Prototheory of Causation." In Collins et al. (2004), 119–137.