Chapter 5: Advertisement for a Sketch of an Outline of a Proto-

Theory of Causation

S. Yablo

!

I. Plato's distinction

A couple of thousand years before Hume made the remark that inspired the counterfactual theory of causation, Plato said something that bears on the principal problems for that theory. The idea will seem at first utterly familiar and of no possible help to anyone, so please bear with me. What Plato said, or had Socrates say, is that a distinction needs to be drawn between "the cause" and "that without which the cause would not be a cause" (Phaedo, 98e).

This sounds like the distinction between causes and enabling conditions: conditions that don't produce the effect themselves but create a context in which something else can do so; conditions in whose absence the something else would not have been effective. And, indeed, that is what Plato seems to have had in mind. Crito offers Socrates a chance to escape from prison. Socrates refuses and sends Crito on his way. The cause of his refusal is his judgment that one should abide by the decision of a legally constituted court. But it is facts about Socrates's body that allow the judgment to be efficacious: "if he had not had this

apparatus of bones and sinews and the rest, he could not follow up on his judgment, but it remains true that it is his judgment on the question that really determines whether he will sit or run" (Taylor 1956, pp. 200-1).

Socrates's bones and sinews are factors such that if you imagine them away, the cause (Socrates's judgment) ceases to be <u>enough</u> for the effect. Are there conditions such that the cause ceases to be <u>required</u> for the effect, if you imagine them away? There seem to be. Consider an example of Hartry Field's.

BOMB: Billy puts a bomb under Suzy's chair; later, Suzy notices the bomb and flees the room; later still, Suzy has a medical checkup (it was already arranged) and receives from her doctor a glowing report.

Field intends this as a counterexample to transitivity, and so it is. The bomb is a cause of the fleeing is a cause of the glowing report; the bomb is not a cause of the glowing report. But it is also an example of Plato's distinction. Were it not for the bomb's presence, the glowing report would not have hinged on Suzy's leaving the room. The bomb does not help Suzy's leaving to suffice for the glowing report; rather it makes Suzy's action important, required, indispensable.

Apparently there are two kinds of factors "without which the cause would not be a cause." On the one hand we have <u>enablers</u>: facts G such that $(OC \& \neg G) >> \neg Oe$. On the other we have what might be called <u>ennoblers</u>: facts G such that $(\neg OC \& G) >> \neg Oe$. Enablers make a dynamic contribution. They help to bring the effect about. What an ennobler contributes is just a raising of status. Suzy's removing herself from the room is elevated from something that just happens to something that <u>had</u> to happen, if Suzy was later going to be healthy.

Plato thinks that factors "without which the cause would not be a cause" are one thing, causes another. He presumably then would say that enablers and ennoblers are not to be regarded as causes.

About enablers, at least, it is not clear we should go along with him. If G is an enabler, then it is a fact in whose absence the effect would not have occurred. And, although there is some dispute about this, most say that that is good enough for being a cause. Enablers are fully-fledged causes; it is just that they are pragmatically counterindicated in some way. Plato's distinction in its cause/enabler form can easily be rejected. Let's suppose to keep things simple that it is rejected.

Consider now ennoblers. An ennobler contributes by closing off potential routes to \underline{e} , viz. all the routes not running through \underline{c} . This if anything <u>hurts</u> \underline{e} 's chances. So there is no question of confusing an ennobler with a cause. Plato's distinction in its

(unintended) cause/ennobler form is real and important. That a potential cause is disarmed may be a factor in the manner of \underline{e} 's occurrence, but it is not a factor in its occurring as such.

II. Preemptive causes

The counterfactual theory as handed down from Hume says that "if the first did not, then the second had not been." For this to be plausible, we would need to add that the first and second both occur, that the first is distinct from the second, and so on. But let us imagine that all of that is somehow taken care of, because our concerns lie elsewhere. Let the "simple" counterfactual theory be just what Hume says:

(CF1) <u>c</u> causes <u>e</u> iff <u>e</u> depends on <u>c</u>, that is, $\neg O(\underline{c}) >> \neg O(\underline{e})$

The simple theory cannot be right, because it ignores the possibility of back-up causes that would spring into action if the real cause failed. This is what Lewis used to call the asymmetric overdetermination problem, and now calls preemption. An example is

DEFLECT: Hit and Miss both roll bowling balls down the lane. Hit's heavier ball deflects Miss's lighter ball en route to the pin. Hit's throw caused the pin to fall. But there is no dependence since if it had not occurred, the effect would

still have happened due to a chain of events initiated by Miss's throw. (Yablo 1986, p. 143)

The problem here is obvious enough that one should probably date the counterfactual <u>theory</u>, as opposed to the immediately withdrawn counterfactual <u>hunch</u>, to the moment when it was first clearly addressed, in Lewis's 1973 paper "Causation."

Here is what Lewis says. Notice something about the chain of events initiated by Miss's throw. The Miss-chain was cut off before the Hit-chain had a chance to reach the pin. It is true that the effect does not depend on the earlier part of Hit's chain. It does, however, depend on the part occurring later, after Miss's chain is dead and buried. And the part after the cut-off point depends in turn on Hit's throw. So the effect depends on something that depends on Hit's throw -- which suggests that instead of (CF1) our analysis should be

(CF2) <u>c</u> causes <u>e</u> iff $\exists \underline{d}_i$ such that $\neg O(\underline{c}) \gg \neg O(\underline{d}_1) \& \dots \& \neg O(\underline{d}_n) \gg \neg O(\underline{e})$

What (CF2) says is that causation need not be direct; it can be indirect, involving dependency chains. If the ancestral of a relation R is written R*, it says that causation is dependence*.

The diagnosis implicit in (CF2) is that preemption arises because we had forgotten about causal chains. I want to suggest an alternative "Platonic" diagnosis. Preemption happens because to take away a cause \underline{c} is, sometimes, to take away more. It is to take away the reason it <u>is</u> a cause. It is to take away factors which, although not themselves causal, contribute to \underline{c} 's causal status by putting \underline{e} in need of \underline{c} . \underline{E} can hardly be expected to follow \underline{c} out of existence, if the reasons for its depending on \underline{c} disappear first!

So, look again at DEFLECT. Quoting a former self: "if in fact Miss's ball never reaches the pin, then that is an important part of the circumstances. Relative to circumstances including the fact that Miss's ball never makes it, what Hit did was necessary for the pin's toppling. If in those circumstances Hit hadn't rolled his ball down the alley, the pin would have remained standing." (Yablo 1986, p. 159) That Miss's ball never touches the pin is a fact that puts the effect in need of Hit's throw. It is a fact "in virtue of which Hit's throw is a cause." The trouble is that it is a fact put in place by the throw itself, hence one that finkishly disappears when the relation is counterfactually tested.

III. Holding fixed

The diagnosis suggests a repair. If preemption is a matter of something finkishly giving way, the obvious thought is: don't <u>let</u>

it give way; hold the grounds of the causal connection fixed. The test of causation in these cases is not whether \underline{e} fails if \underline{c} does, but whether \underline{e} fails if \underline{c} fails \underline{with} the right things held fixed.

By "dependence modulo G" I will mean dependence with G held fixed. This event depends modulo G on that one iff had that one failed to occur in G-type circumstances, this one would have failed to occur as well. Letting '>>_G' stand for dependence modulo G, the suggestion is that

(CF3) <u>c</u> causes <u>e</u> iff: for some appropriate G, $\neg O(\underline{c}) \gg_{G} \neg O(\underline{e})$,

Actually, of course, this is only an analysis-schema. An analysis would require a clear, non-causation-presupposing, statement of what makes for an appropriate G.

Well, what does make for an appropriate G? Certainly G should ennoble \underline{c} . But all we have said about ennoblers is that they are conditions G such that \underline{e} depends holding-G-fixed on \underline{c} . As you might guess, and as will be discussed below, this purely formal requirement can be met by logical trickery almost whatever \underline{c} and \underline{e} may be.

Thus where the standard counterfactual theory undergenerates -the events that depend on \underline{c} (or depend* on \underline{c}) are not all the events it causes -- the present theory has, or is in danger of

having, the opposite problem. It may well be that \underline{e} depends on each of its causes modulo a suitably chosen G. But this is true also of events that do not cause e.

IV. Triviality/polarity

There are actually two worries here, one building on the other. The first is a worry about trivialization: everything depends on everything modulo a silly enough G. This is illustrated by

JUMP: Suppose that \underline{e} is Bob Beamon's jumping 29'2 ½" at the 1968 Olympics in Mexico City. And let \underline{c} be the burning out of a meteor many light years away. It is not hard to find facts modulo which the jump depends on the burn-out. First choose an event on which the jump depends pure and simple -- say, Beamon's tying his shoes. Holding it fixed that <u>the tying</u> <u>occurs only if the burn-out does</u>, without the burn-out Beamon does not make the jump.

Unless some sort of restriction is put on admissible G's, the requirement of dependence modulo G is a trivial one satisfied by any pair of events you like. One could stipulate that G should not be too cockamamie or too ad hoc or too cooked up for the occasion. But that is hopeless. It is not just that "too ad hoc" is so vague. Suppose that a standard of naturalness is somehow agreed

on. No matter how high the standard is set, there will be Gdependence without causation.

Consider again BOMB. Certainly the doctor's glowing report does not depend simpliciter on Billy's planting the bomb. But it does depend on it modulo the fact that <u>Suzy's chair explodes</u>. Holding the explosion fixed, Suzy would not have been healthy unless she had moved away, which she would not have had she not noticed the bomb, which would not have been there to notice had it not been put there by Billy. It seems on the face of it insane to credit Suzy's good health to the bomb; she is healthy despite the bomb, not because of it! And yet her health depends on the bomb modulo a pretty natural fact. Call that the polarity problem.

V. Stockholm Syndrome

Preemptive causes make themselves indispensable. They create the conditions given which the effect would not have occurred without them. But there is more than one way of doing that. The <u>normal</u> way is to produce the effect yourself, thereby preventing other wouldbe causes from doing the job instead. The effect needs \underline{c} modulo the fact G that other avenues to the effect are closed off.

But if you look at our basic condition -- the condition of dependence modulo G -- you can see that it supports an almost opposite scenario. \underline{E} was going to happen anyway, when \underline{c} comes along to threaten it: to put its existence in jeopardy. Of course,

putting the effect in jeopardy is not all \underline{c} does, or it would not even resemble a cause. It also rescues \underline{e} from the jeopardy. \underline{C} threatens \underline{e} with one hand, and saves it with the other. The effect needs \underline{c} to counter the threat G that \underline{c} itself has launched.

Now, should \underline{e} be grateful to \underline{c} for blocking with one hand a threat it launches with the other? Of course not. There is a word for that kind of inappropriate gratitude. You might remember it if I quote from a website on the topic: "In the summer of 1973, four hostages were taken in a botched bank robbery at Kreditbanken in Stockholm, Sweden. At the end of their captivity, six days later, they actively resisted rescue. They refused to testify against their captors, raised money for their legal defense, and according to some reports one of the hostages eventually became engaged to one of her jailed captors." Stockholm Syndrome is the gratitude hostages feel toward captors who help them with problems brought on by the captivity. To give the present sort of \underline{c} causal credit would be the metaphysical equivalent of Stockholm Syndrome.

It is important to be clear about what is being rejected here. There is nothing wrong with gratitude for actions taken against a threat that has already been launched: not even if the action is taken by the one who launched it. If your kidnapper takes pity on you and gets you a Mars bar, there is no requirement of flinging it back in his face. But suppose your kidnapper says, "I appreciate that you are grateful to me for various particular acts

of mercy. But still I am hurt. Where is the thanks I get for the action that occasioned these mercies, that is, the kidnapping?" That remark you <u>should</u> fling back in his face. I draw the following

<u>Moral</u>: Dependence modulo G does not make for causation if (i) G is a threat to \underline{e} that, although (ii) countered by \underline{c} , was also (iii) launched by c.

With this in mind, let's go back to the BOMB example. There is nothing wrong with thanking the bomb for tipping you off to its presence, given that it has already been planted. But what we are talking about here is gratitude toward the planting itself. BOMB, then, is an example of Stockholm Syndrome. What may be less obvious is that JUMP is an example as well.

That there will be no shoe-tying unless the meteor burns out makes <u>e</u> vulnerable from a new and unexpected direction; events that would cancel the burn-out are now being put in a position to cancel the jump too. (That's (i).) There would been no such fact as <u>shoe-tying only if burn-out</u>, had the meteor not in fact burned out. (That's (iii).) It is the burn-out, finally, that stops this fact from carrying out its threat against the effect. (That's (ii).)

VI. Artificial needs

If <u>e</u> depends on <u>c</u> holding G fixed, let us say that <u>G puts e in</u> <u>need of c</u>. Why is this not enough for causation? The answer is that some needs are trumped up or artificial. This shows up in the fact that among <u>e's other</u> needs are some that would, but for <u>c</u>, have been <u>all</u> its needs. Or, to look at it from the point of view of the fallback scenario -- the closest scenario where <u>c</u> does not occur -- <u>c</u> is able to meet a need only by making the effect needier than it had to be, indeed, needier than it <u>would</u> have been had c failed to occur.

Say that \underline{e} is Beamon's big jump, and \underline{c} is the burning out of that meteor. What are the effect's needs in the fallback scenario? What would the jump have depended on had the meteor not burned out? It would have depended on Beamon's tying his shoes; on various earlier jumps whereby he won a place in the finals; on Mexico's bid for the 1968 Olympics; and so on. <u>Bringing in the burn-out does not diminish these needs one iota</u>. Everything that had to happen before, still has to happen with the meteor burning out. This is why the burn-out's role is artificial. It is strictly additional to events that meet the effect's needs all by themselves in its absence.

Now let's try to make this a teeny bit precise. History let's suppose has a branching time structure. There is the trajectory actually taken through logical space, and the various branchings-

off corresponding to other ways things could have developed. One branch in particular corresponds to the way things would have developed if \underline{c} had not occurred. By the <u>fallback scenario</u> let's mean what happens after the branching-off point on that alternative branch. By the <u>actual scenario</u> let's mean what actually <u>does</u> happen after the branching-off point. The effect's fallback needs are the events it depends on in the fallback scenario, that is, the events it would have depended on, had \underline{c} not occurred.¹ Recalling our subscripting conventions from above, this can be written

$$FAN = \{\underline{x} \mid \neg O\underline{x} \gg_{F} \neg O\underline{e}\},\$$

where 'F' is short for ' $\neg O\underline{c}$.'² The effect's actual needs (for a given choice of G) are the events it depends on modulo G in the actual scenario. This can be written

```
GAN = \{x \mid \neg Ox >>_G \neg Oe\}.
```

The need for <u>c</u> is artificial iff GAN covers FAN with <u>c</u> to spare; or, taking the perspective of the fallback scenario, FAN is identical to a subset of GAN not including <u>c</u>. (I will write this FAN = GAN^{-} .) The point either way is that <u>c</u> speaks to a need that

is piled arbitrarily on top of what would, in \underline{c} 's absence, have been all the needs.

VII. Counterparts

Imagine that Beamon as a child was initially attracted to chess rather than long jump. He was waiting to sign up for chess club when someone threw a rock at him. It was because of the rock incident that he wound up in track. And now here is the interesting part: the rock came from the burnt-out meteor. If not for the burn-out, it would not have been <u>that</u> rock-throwing the effect depended on but a related one (in which the bully hurled a different rock). Since the rock-throwing that <u>e</u> would have needed is a different event from the one it does need, it would seem that GAN does not cover FAN at all, let alone with c to spare.

I answer that different <u>event</u> does not have to mean different <u>need</u>. One event can meet the same need as another, as that need manifests itself in their respective scenarios.⁴ Artificiality is still a matter of the effect's actual needs subsuming its fallback needs, so long as we understand this in the following way. Suppose that \underline{x} is an event needed in the fallback scenario; one finds in GAN, not perhaps that very event, but <u>an</u> event meeting the same need (henceforth, a counterpart of \underline{x}). This complicates things a little, but not much. Where earlier we required FAN to be identical to a subset of GAN not including \underline{c} , now we ask only that

it coincide with such a subset, where sets coincide iff their members are counterparts.⁵ (I will write this FAN \approx GAN⁻.)

When do events speak in their respective scenarios to the same need? The idea is this. Needs that \underline{e} would have had in \underline{c} 's absence can be paired off with actual needs in ways that preserve salient features of the case: energy expended, distance traveled, time taken, place in the larger structure of needs. One wants to preserve as many of these features as possible, while finding matches for the largest number of needs. One asks: how much of the fallback structure is embeddable in the actual one? What is the maximal isomorphic embedding? Events speak to the same need if they are linked by this embedding.

VIII. De facto dependence

It is not enough for causation that a G can be found that puts \underline{e} in need of \underline{c} . Causes must meet <u>real</u> needs, and the need met by \underline{c} might be trumped up or artificial. A fact G makes the need for \underline{c} artificial iff it assigns other needs that would, but for \underline{c} , have been all of e's needs.

An issue I have finessed until now is how the first G -- the one that puts \underline{e} in need of \underline{c} - lines up with the second one -- the one that makes the need artificial. Suppose we say of the second G that it "enfeebles" \underline{c} , as we said of the first that it ennobles

it. Does it suffice for causation that an ennobler G exists that is not <u>itself</u> an enfeebler?

No, for there is almost always a G like that, namely, the material biconditional <u>e occurs <-> c occurs</u>. That this ennobles <u>c</u> should be clear. That it does not enfeeble <u>c</u> can be seen as follows. (1) GAN is limited to events <u>x</u> on which <u>c</u> counterfactually depends. (If $\neg O\underline{x} \gg O\underline{c}$, ⁶ then $(O\underline{c} <-> O\underline{e}) \gg (\neg O\underline{x} \approx O\underline{c})$; so by the export-import law, $(\neg O\underline{x} \& (O\underline{c} <-> O\underline{e})) \gg O\underline{c}$; so $(\neg O\underline{x} \& (O\underline{c} <-> O\underline{e})) \gg O\underline{e}$ iff $(\neg O\underline{x} \& (O\underline{c} <-> O\underline{e}) \& O\underline{c}) \gg O\underline{e}$; so $(\neg O\underline{x} \& (O\underline{c} <-> O\underline{e})) \gg O\underline{e}$; so $(\neg O\underline{x} \& (O\underline{c} <-> O\underline{e})) \gg O\underline{e}$; so \underline{x} is not in GAN.) (2) FAN is almost certainly <u>not</u> limited to events on which <u>c</u> counterfactually depends. That <u>c</u> fails to depend on <u>x</u> has no tendency at all to suggest that <u>e</u> would not have depended on <u>x</u> in <u>c</u>'s absence. (1) and (2) make it unlikely that GAN includes FAN, or hence that G enfeebles <u>c</u>.

Where does this leave us? It is not enough for causation that an ennobler G can be found that is not itself an enfeebler. It is, I suggest, enough that an ennobler can be found such that <u>no</u> comparably natural enfeeblers exist. And so I propose a definition

(DD) one event <u>de facto depends</u> on another iff some G putting the first in need of the second is more natural than any H that makes the need artificial,

and I make the following claim

(CF3) c is a cause of e iff e de facto depends on c.

It is understood that \underline{c} and \underline{e} both occur, that they are suitably distinct, and that various unnamed other conditions are met; I have in mind the same sorts of extra conditions as the counterfactual theorist uses. Sometimes (CF3) will be written (DF) to emphasize that it relies on a new type of dependence, albeit one defined in terms of counterfactual dependence.

You might have expected me to say that \underline{c} is a cause iff \underline{e} either depends counterfactually on \underline{c} or, failing that, de facto depends on it. That formulation is fine but it is equivalent to what I did say, for de facto dependence has ordinary counterfactual dependence as a special case. If \underline{e} counterfactually depends on \underline{c} , then it depends on \underline{c} modulo the null condition. The null condition is our ennobler and what needs to be shown is that there are no comparably natural enfeeblers. But there cannot be enfeeblers at all, for enfeeblers presuppose fallback needs -- events that \underline{e} depends on in \underline{c} 's absence -- and \underline{e} does not even occur in \underline{c} 's absence.

IX. Triviality and polarity

One worry we had is that even if \underline{c} and \underline{e} are completely unrelated, still \underline{e} is put in need of \underline{c} by the fact that \underline{k} occurs only if \underline{c} occurs, where \underline{k} is an event on which the effect counterfactually depends.

I say that while this is true, the victory is short-lived, because the very fact of unrelatedness means that it will be easy to find an H making the need artificial. Usually we can let H be the null condition. That is, \underline{e} counterfactually depends outright (holding nothing fixed) on events that would have been enough in \underline{c} 's absence. This is just what we would expect if \underline{c} has causally speaking nothing to do with \underline{e} . Beamon's jump depends on all the same things if the burn-out occurs as it would have depended on absent the burn-out.

The need for \underline{c} is artificial iff it is over and above what would, but for \underline{c} , have been all the needs. An equivalent and perhaps clearer way of putting it is that c must either <u>meet</u> a fallback need -- which it does if for some \underline{f} in FAN, \underline{c} meets the same need as \underline{f} -- or <u>cancel</u> one -- which it does if for some \underline{f} in FAN, no actual event meets the same need as \underline{f} . The need for \underline{c} is artificial iff \underline{c} fails to address any fallback needs, meaning that it neither meets any fallback needs nor cancels any.

I take it as given that Billy's planting of the bomb does not meet any fallback needs. The question is whether it cancels any.

Suppose that Suzy needs to stay hydrated, or she becomes very sick. She has set her Palm Pilot to remind her at noon to act on this need. The fallback scenario has her sitting quietly in her chair at noon. She has a drink of water, water being the one hydrous stuff available in the room. The actual scenario has Suzy catching her breath on the sidewalk when her Palm Pilot beeps. She eats some Italian ice, that being the one hydrous stuff available on the sidewalk. Any isomorphism worth its salt is going to associate these two events. The drinking and eating are counterparts; they speak to the same need. One imagines that the same can be done for all of the effect's fallback needs. Anything the glowing report needed absent the bomb, it still needs. The reason Billy's action is not a cause is that it fails to address any fallback needs.

Suppose that I am wrong about that. Suppose the effect's fallback needs are <u>not</u> all preserved into the actual situation; or suppose they are all preserved but one maps to the planting of the bomb. Then, I claim, the planting starts to look like a cause.

<u>Case 1</u>: There is an <u>f</u> in FAN such that Billy's action meets the same need as <u>f</u>.

Suzy needs exercise or she becomes very sick. She has set her Palm Pilot to remind her to exercise at 11:45. As things turn out,

she doesn't hear the beeping because she has just spotted a bomb under her chair. Running from the bomb gives her the needed exercise and so saves her health. If that is how it goes, then Billy's planting the bomb meets the same need as would have been met by Suzy's setting her Palm Pilot. And now we are inclined to reason as follows. Billy's planting the bomb meets the need for an exercise-reminder; the need was not artificial because it would have been there bomb or not; so there is no objection to treating what Billy did as a cause.

<u>Case 2</u> There is an \underline{f} in FAN such that no actual event meets the same need as f.

Billy's planting the bomb does not in fact meet the same need as Suzy's setting her Palm Pilot. The Palm Pilot, if she had heard it, would have led Suzy to do push-ups, thus exercising her muscles. The bomb leads her instead to run, thus exercising her heart and lungs. These are entirely different forms of exercise. Either one of them would have stopped Suzy from getting sick, but the similarity ends there. Now we are inclined to reason as follows. The effect originally had need of <u>muscle</u> exercise, that being the only kind of exercise possible in the room. It is <u>relieved</u> of that need by Billy's planting of the bomb; for Suzy now runs, thus exercising her heart and lungs. So there is no

objection to treating what Billy did as a cause. (Analogy: You have a flat tire and need a jack to get back on the road. I can help you either by meeting that need, or by relieving you of it. I do the first if I provide you a jack. I do the second if I bend over and lift the car myself.)

X Preemption

I say that effects really do depend on their preemptive causes. There is no counterfactual dependence, because the causality rests on a fact G; and had \underline{c} not occurred, that fact would not have obtained. But we can restore the dependence by holding G fixed. I don't know how to argue for this except by going through a bunch of examples.

Recall DEFLECT. Certainly the effect is put in need of Hit's throw by the fact G that Miss's ball never gets close to the pin. It might be thought, though, that the need was artificial.

The effect's fallback needs are (let's say) for Miss's throw, her ball's rolling down the aisle, and her ball's hitting the pin. These needs would seem to recur in the actual situation as needs for Hit's throw, his ball's rolling down the aisle, his ball's hitting the pin. If that is how things line up, then Hit's throw meets the same need as was met in the fallback scenario by Miss's throw; and so the need it meets is not artificial.

Suppose on the other hand that the fallback needs are held <u>not</u> to recur in the actual situation. Then artificiality is averted through the canceling of needs rather than the meeting of them. These are intuitive considerations but they suggest that a fact making the need for Hit's throw artificial will not be easy to find. I do not doubt that you could construct one by brute force, but a brute force H will not be as natural as our existing G, the fact that Miss's ball never gets close.

A tradition has arisen of treating early and late preemption as very different affairs. But this is for theoretical reasons to do with Lewis's ancestral maneuver, which works for early preemption but not late; intuitively the two sorts of preemption seem much on a par. The de facto theory agrees with intuition here. Consider

DIRECT: Hit and Miss both roll balls down the lane. The balls do not come into contact. Hit's ball knocks the pin into the gutter. A moment later, Miss's ball reaches the spot where the pin formerly stood.

Once again, it is part of the circumstances that Miss's ball never gets close to the pin. That no other ball gets close puts the effect in need of Hit's throw. It is true that some H might expose the need as artificial. But such an H would have to be

constructed by brute force. There is no more reason to expect a natural enfeebler in this case than the previous one.⁷

XI. Overdetermination

Overdetermination occurs when an effect \underline{e} depends on two events taken together without depending on either taken alone; and (what distinguishes it from preemption) neither can lay claim to being more of a cause than the other. Consider

TOGETHER: Knock and Smack roll their balls at the same time; the balls hit the pin together and it falls over; either ball alone would have been enough.

It is not hard to find suitable G's. The effect depends on Knock's throw holding fixed the fact G_k that Smack's ball does not hit the pin unaccompanied, that is, unless another ball also hits. And it depends on Smack's throw holding fixed the fact G_s that Knock's ball does not hit the pin unaccompanied.

It is not hard to find suitable H's either; indeed we have already found them. G_k makes the need for Smack's throw artificial, and G_s does the same for Knock's. To see why, suppose that Knock had not thrown. The effect would have depended on Smack's throw, the forward motion of his ball, and the like. These

events are still needed in the actual situation, if we hold fixed the fact G_s that Knock's ball does not hit alone.

Assuming that these are the most natural cause-makers and breakers to be had, does the de facto theory call Knock's throw (e.g.) a cause? Is the effect put in need of it by a fact more natural than any fact making the need artificial?

That depends. One reading of "more natural" is <u>strictly</u> more natural. If that is what is meant, then neither throw is a cause; each prima facie connection is broken by a fact exactly as natural as the one that established it. But the phrase could also be taken weakly, to mean "at least as natural as." If, as claimed, the makers and breakers are the same, then the weak reading makes both throws out to be causes. True, each occurs under conditions given which the effect takes no notice of it; but then each also occurs under conditions no less natural given which the effect needs it. Ties go to the runner on the weak reading, so we have two bona fide causes. Our uncertainty about overdeterminers reflects indecision about what to mean by "more". (This is intended less as an explanation of the uncertainty than a rational reconstruction of it.)

XII. Asymmetry

Suppose that \underline{c} affects not whether \underline{e} occurs but only when it occurs. Could that be enough to make \underline{c} a cause? An example is given by Jonathan Bennett.

RAINDELAY: "There was heavy rain in April and electrical storms in the following two months; and in June the lightning took hold and started a forest fire. If it hadn't been for the heavy rain in April, the forest would have caught fire in May." (Bennett 1987, p. 373)

Bennett says that "no theory should persuade us that delaying a forest's burning for a month (or indeed a minute) is causing a forest fire...." And then he points out something interesting. "Although you cannot cause a fire by delaying something's burning, you can cause a fire by hastening something's burning." (<u>ibid</u>.) So, consider

LIGHTNING: There are no rains in April. The fire happens in May due to May lightning, rather than in June due to the lightning that strikes then. The lightning is a cause of the fire even though the fire would still have occurred without it. That the time of occurrence would have been later rather than earlier seems to make all the difference.

Bennett's examples raise two problems for standard counterfactual accounts. One is that they cannot explain the <u>asymmetry</u>, that is, why hasteners seem more like causes than delayers. Also, though, they have trouble explaining <u>why there</u> <u>should be causation here at all</u>. I assume with Bennett that hasteners bring it about that the very same event occurs earlier than it would have. If in fact the fire would still have occurred without the lightning, how can the lightning be regarded as a cause?

The form of that question ought to seem pretty familiar. It is the standard preemption question. How can \underline{c} be a cause, when the effect would have occurred without it thanks to \underline{c} ' waiting in the wings? The answer is the same as always: It is a cause because the effect depends on it modulo a certain pretty natural fact and nothing that natural exposes the dependence as fraudulent. It is a part of the circumstances that <u>the woods do not catch fire in June (or later</u>). Holding that fixed, without the May lightning there would not have been a fire. The May lightning causes the fire because the fire depends on it holding fixed that May is its last opportunity.

But there is an obvious objection. The effect also fails to occur <u>before</u> a certain time, and this would seem to obliterate the intended asymmetry. Holding fixed the lack of a fire before June,

if not for April's rain there would not have been a fire at all. June was the window of possibility, and it was the rain that kept the forest going until it opened.

The difference between rain and lightning is not that the first meets no need. It has to do with the kind of need. Suppose the rain had not fallen, so that the forest burned in May. Then the things that were done to preserve it from May until June would not have been required. (The loggers wouldn't have had to go on strike, the rangers wouldn't have had to apply the flame retardant, and so on.) That the rain introduces new needs would not be a problem if it addressed some old ones. But it doesn't. The things that would have been needed for the May fire, had the rain not fallen, continue to be needed as conditions of the June fire. (A landslide late in April threatens to bury the forest under rubble; the June fire needs it to change course just as much as the May fire would have.)

Now we see why the rain makes a bad cause. It piles on new needs without canceling any old ones. The lightning, by contrast, cancels a whole month of old needs. The pattern here is typical of the genre. Just by their definition, hasteners are liable to speak to fallback needs; they reduce the time period over which the effect is in jeopardy and so cancel any needs pertaining to the period that is chopped off. Just by their definition, delayers often bring about a situation in which the effect needs more than

it would have had the delaying event not occurred. The effect is in jeopardy for longer and has needs pertaining to the extra time. This is why hasteners tend to be causes and delayers tend not to be.

XIII. The hastener theory

I have treated hastening as a special case of preemption. One might try the reverse, assimilating preempters to hasteners (Paul 1998). A cause is an event in whose absence <u>e</u> would have not occurred, or would not at any rate have occurred as early as it did. If we count never occurring as the limiting case of delay, then the claim is that causes are hasteners, that is, events in whose absence the effect would have been delayed. One problem for this view is that hasteners are not always causes. Here is an example due to Hugh Rice (1999, p. 160):

REFLEXES: Slow Joe and Quick-Draw McGraw are shooting at Billy the Kid. Joe fires first, but since his gun fires slower-moving bullets, it is not too late for McGraw (if he fires) to cause the death. And so it happens. "McGraw (blest...with super fast reflexes) was aware of Joe's firing and as a result (wishing to have the glory of killing Billy for himself) fired a little earlier than he would otherwise have done.... It seems that McGraw's firing was a cause of \underline{e} ,

but that Joe's firing was not." Both shots hasten the death. So both count on the hastener theory as causes. Intuitively, however, it is McGraw's shot that kills Billy.

What does the de facto theory say about this? It is not hard to find a G modulo which the death depends on McGraw's shot. As the situation in fact develops, Joe's bullet never comes into contact with Billy (it passes untouched through the hole left by McGraw's bullet). Holding that fixed, Billy's death would not have occurred were it not for McGraw. This same G also enfeebles Joe's shot. Had Joe not fired, Billy's death would have depended on McGraw's shot, the motion of his bullet, and so on. Those are its fallback needs. The death's actual needs are the events on which it depends holding fixed that Joe's bullet never made contact. Prima facie it would seem that the death's fallback needs are all preserved into the actual scenario: anything the effect depended on absent Joe's shot, it continues to depend on given that Joe's bullet doesn't hit anything.

I said that hasteners tend to reduce needs pertaining to the time period over which the effect is no longer in jeopardy. That assumes, however, that the counterpart relation puts a lot of emphasis on temporal as opposed to other factors. Oftentimes other factors will seem just as important, or more important. Suppose that by kicking a bowling ball already en route to the pin, I get

it to arrive more quickly. Ordinarily my kick would count as a cause. This time, though, the main threat to the ball's forward motion is from equally spaced gates that open and shut according to a complicated pattern. The effect occurs only if the ball makes it through each of the gates. Then we might feel that the effect's needs are better conceptualized in terms of number of gates than number of seconds. To the extent that kicking the ball leaves its chances with the gates unchanged, the "need" it meets comes to seem artificial. Certainly the kick seems like less of a cause when it is stipulated that the obstacles are spatially distributed rather than temporally.

I said that delayers often bring about a situation in which the effect needs strictly more than it would have, had the delaying event not occurred. The effect is in jeopardy for longer and has needs pertaining to the extra time. But again, this is only a trend, not a strict rule. Sometimes by putting an effect off for a bit we can cut down on other and more important needs. Consider a variant of REFLEXES. McGraw is standing further from Billy than Slow Joe. When Joe sees that McGraw has fired, he fires his slower bullet on a trajectory that has it deflecting McGraw's bullet off to the side before reaching Billy. Joe's firing makes the effect happen later than it would have, but it is still a cause. Counterparthood is judged not in respect of time but dependency relations; Joe's firing meets the need that McGraw's

would have met or, on an alternative accounting, cancels it. It is Joe's shot that kills Billy, despite the fact that Billy lives a little longer because of it.

XIV. Trumping preemption

A second recent response to the preemption problem focuses on events causally intermediate between \underline{c} and \underline{e} . It exploits the fact that, in all the usual cases, \underline{e} would have depended on events other than those actual intermediaries had \underline{c} failed to occur (Ganeri, Noordhof, and Ramachandran 1998). A third focuses on the manner in which the effect occurs, if caused by something other than \underline{c} . There is nothing in the nature of preemption, though, that requires intermediate events, or that the effect's characteristics should vary according to its cause.

SPELL: Imagine that it is a law of magic that the first spell cast on a given day [matches] the enchantment that midnight. Suppose that at noon Merlin casts a spell (the first that day) to turn the prince into a frog, that at 6:00pm Morgana casts a spell (the only other that day) to turn the prince into a frog, and that at midnight the prince becomes a frog. Clearly, Merlin's spell...is a cause of the prince's becoming a frog and Morgana's is not, because the laws say that the first spells are the consequential ones. Nevertheless, there

is no counterfactual dependence of the prince's becoming a frog on Merlin's spell, because Morgana's spell is a dependency-breaking backup. Further, there is neither a failure of intermediary events along the Morgana process (we may dramatize this by stipulating that spells work directly, without any intermediaries), nor any would-be difference in time or manner of the effect absent Merlin's spell...thus nothing remains by which extent [counterfactual accounts of causation] might distinguish Merlin's spell from Morgana's in causal status. (Schaffer, chapter 2 in this volume, p. REF)

What does our sketch of a proto-theory say about this case? First we should look for a G such that the effect depends modulo G on Merlin's spell. How about the fact that no one casts a spell before Merlin does? Holding that fixed, there would have been no transformation had Merlin not cast his spell. Perhaps a no less natural H can be found that enfeebles Merlin's spell; I have not been able to think of one. It is perhaps enough to show that, unlike the other approaches mentioned, the de facto dependence account is not at an absolute loss here.

XV. Switching

A switch is an event that changes the route taken to the effect. It may not be obvious how switching so described goes beyond standard preemption, but consider an example.

YANK: A trolley is bearing down on a stalled automobile. The car lies 110 yards ahead on the track -- or rather tracks, for just ahead the track splits into two 100-yard subtracks that reconverge 10 yards short of the car. Which subtrack the trolley takes is controlled by the position of a switch. With the switch in its present position, the trolley will reach the car via subtrack U (for unoccupied). But Suzy gives the switch a yank so that the trolley is diverted to subtrack O (for occupied). It takes subtrack O to the reconvergence point and then crashes into the car.

Certainly the crash does not counterfactually depend on the yank; had Suzy left the switch alone, the trolley would have taken subtrack U to the car, and the crash would have occurred as ever. Thus the simple counterfactual theory (CF1) does not classify the yank as a cause. The ancestralized theory (CF2) sees things differently; the effect depends on events that depend on the yank -- the trolley's regaining the main line from track O, for instance -- so what Suzy did was a cause. (The verdict does not

change if track 0 was mined; Suzy was hoping to get the trolley blown up, and would have succeeded had not the bomb squad arrived.)

What does the de facto theory say? There is no trouble finding a G such that the crash depends modulo G on the yank. Holding fixed that subtrack U is untraveled, had the switch not been pulled there would have been no way forward; the trolley would, let's assume, have derailed. The worry is that some comparably natural H makes the need artificial. And, indeed, the null fact makes it artificial. Here in the actual scenario, the effect has need of 100 one-yard motions down track O. Had the yank not occurred, its needs would have been for 100 one-yard motions down track U. Because the yank lies apart from what might as well have been all the effect's needs, the de facto theory does not call it a cause.

The de facto theory lets the yank be a cause iff it either <u>meets</u> a fallback need or <u>cancels</u> one. As the case was first stated, it does neither thing, but suppose we tweak it a little. Suppose that O is shorter, or that U was disconnected when Suzy pulled the switch. Then there are needs the effect would have had which the yank does away with, and so the role it plays is not entirely artificial. Alternatively, suppose the switch operates not by rearranging the tracks, but by physically grabbing hold of the train and forcing it away from U and down O. Then the yank does meet a fallback need, the one that would in its absence have been

met by the train's continued momentum. The door is thus open to the yank's being classified as a cause.

This is a good a place to acknowledge that although technically, everything that e would have depended on counts as a fallback need,⁸ in practice not all such needs are taken equally seriously. Suppose that track U has been disconnected for years, and heroic efforts are required to fix it. That it makes those efforts unnecessary earns the yank causal credit. But what if the track is constantly reversing itself; it is part of U's design to connect when it senses an approaching trolley and disconnect when the trolley is safely past. Then the need that gets cancelled may be considered too slight to protect the yank from charges of artificiality. I have no criterion to offer of when a fallback need is sufficiently serious that c can escape artificiality by canceling it. But two relevant questions are these: Were the effect to fail, what are the chances of its failing for lack of x? And how counterfactually remote are the scenarios where x is the culprit? A fallback need may not count for much if it is the last thing one would think of as the reason why e would fail.

Some have said that an event that makes "minor" changes in the process leading to \underline{e} is not its cause, while an event that makes "major" changes is one. Our theory agrees, if "minor" changes are changes whereby all the same needs have to be met. Consider in

this connection an example of Ned Hall's (chapter 7 in this volume, p. REF):

THE KISS. One day, [Billy and Suzy] meet for coffee. Instead of greeting Billy with her usual formal handshake, however, Suzy embraces him and kisses him passionately, confessing that she is in love with him. Billy is thrilled -- for he has long been secretly in love with Suzy, as well. Much later, as he is giddily walking home, he whistles a certain tune. What would have happened had she not kissed him? Well, they would have had their usual pleasant coffee together, and afterward Billy would have taken care of various errands, and it just so happens that in one of the stores he would have visited, he would have heard that very tune, and it would have stuck in his head, and consequently he would have whistled it on his way home But even though there is the failure of counterfactual dependence typical of switching cases (if Suzy hadn't kissed Billy, he still would have whistled), there is of course no question whatsoever that as things stand, the kiss is among the causes of the whistling.

That seems right: the kiss is among the causes of the whistling. But the example is not really typical of switching cases, or at least, it is missing features present in "pure" cases like YANK.

The effect's fallback needs (its needs absent the kiss) are heavily weighted toward the period after Billy leaves the coffee shop. They include, for instance, Billy's deciding to drop into that particular store, the store's staying open until he arrives, the playing of that particular tune, and so on. It is because Suzy's kiss relieves the effect of this heavy burden of lateafternoon needs that we are ready to accept it as a cause.⁹

Notes for Chapter 5

(Yablo, "Advertisement for a Sketch of an Outline of a Proto-Theory of Causation")

1. If it seems odd to think of events as needs, remember that "need" can mean thing that is needed. ("The dogsled was piled high with our winter needs.") Needs in the ordinary sense do not exist in our system. Their work is done by events considered under a soon to be introduced counterpart relation, the relation of meeting-the-same-need-as.

2. According to the export-import law for counterfactuals, A >> (B >> C) is equivalent to (A & B) >> C. This implies that $(\neg O\underline{x} \& \neg O\underline{c}) >> \neg O\underline{e}$, the membership condition for FAN, is equivalent to $\neg O\underline{c} >> (\neg O\underline{x} >> \neg O\underline{e})$, which says that <u>e</u> would have depended on <u>x</u> had <u>c</u> not occurred. I assume that the law is close enough to

correct for our purposes, or at least that the indicated consequence is close enough to correct.

3. FAN and GAN are to be understood as limited to events occurring <u>after</u> the point at which the actual world and the nearest c-less world begin to diverge.

4. Also, same event does not have to mean same need. An event that meets one need here might meet another there, or it might meet no need at all.

5. I will be taking counterparthood to be symmetric and one-one. But there might be reasons for relaxing these requirements. Take first symmetry. There might be an \underline{x} in FAN whose closest actual correspondent meets, not the same need as \underline{x} , but a "bigger" need: one with the need met by \underline{x} as a part. This closest actual correspondent ought to qualify as a counterpart of \underline{x} . So, the argument goes, counterparts should be events meeting <u>at least</u> the same need, which makes counterparthood asymmetric. There is a similar worry about the one-one requirement. It might take a pair of events to meet the need \underline{x} meets all by itself in the fallback scenario; or vice versa. I propose to ignore these complexities.

6. I get from $\neg(\neg O\underline{x} \gg \neg O\underline{c})$ to $\neg O\underline{x} \gg O\underline{c}$ by conditional excluded middle. CEM is generally controversial, but seems in the present context harmless; we are not trying to show that $\underline{Oc} \iff O\underline{c}$ is <u>bound</u> to ennoble \underline{c} without enfeebling it, but just that this is the likely outcome.

7. What if we change the example so that Miss's ball <u>does</u> hit the pin, after it has been knocked down? Then G should be this: Miss's ball never gets close to the pin when it is in an upright position, i.e. when it is in a condition to be toppled. Holding fixed that Miss's ball never approaches the pin at any <u>relevant</u> time, it remains the case that without Hit's throw, the pin would not have been knocked over.

8. Remember that attention is limited to events occurring <u>after</u> the "branch-point": the point at which the nearest <u>c</u>-less world begins to depart from actuality.

9. I was helped by David Christensen, Cian Dorr, Christopher Hitchcock, Igal Kvart, Krista Lawlor, Laurie Paul, Judea Pearl, Ana Carolina Sartorio, and (especially) Ned Hall, who got me reinterested in these topics. I am grateful to audiences at the University of Toronto, University of Vermont, UCLA, Stanford University, and Princeton University. This paper is based in part

on "De Facto Dependence," <u>Journal of Philosophy</u> (forthcoming). The theory presented here is probably closest to the causal-graph theory of Judea Pearl (REF) and Christopher Hitchcock (REF).