

Logic and Metaphysics

CAUSATION

DAVID PAPINEAU

SEMESTER 1 2000-2001 Mondays 11 am 2B08 Strand Bldg KCL

The Regularity Theory

Hume

Our idea of causation is not just of spatio-temporal contiguity. It involves an element of necessitation. Whence this idea? We form it when we find two kinds of events, A and B, constantly conjoined. (This still doesn't account for the idea of necessitation, since constant conjunction is just A and B happening to go together, repeatedly. Hume gave a psychological explanation for our idea of necessitation; causation in the objects is nothing more than constant conjunction.)

Laws and Accidents

Most modern philosophers have focused on the idea of constant conjunction, and forgotten about any extra 'necessitation'. But there are constant conjunctions (true universal gens) where A and B are intuitively quite unconnected causally. We want to distinguish the lawful regularities from the accidents. There are two general strategies for dealing with this problem:

- (i) Humeans stick to the idea that laws just state constant conjunction, and try to explain why some such statements are better than others (they fit into theory, are inductively supportable);
- (ii) Non-humeans (Armstrong) say that laws state relationships of necessitation, which claim more than constant conjunction.

Mackie and INUS Conditions

What shape of regularity is required for A to cause B? We don't seem to want A sufficient (by law) for B -- short circuits aren't always followed by fires. But nor do we want A necessary for B -- we can get fires without short circuits. (Don't confuse A necessitates B, with A is necessary for B (the latter just means $(x)(-Ax \rightarrow -Bx)$)).

Mackie said that A is a cause of B just in case A is an INUS condition of B: A is an Insufficient but Necessary part of a set of conditions which are together Unnecessary but Sufficient.

The idea is that A&M&N, say, Suffice for B; but A alone doesn't -- it's Insufficient; and M&N alone don't

suffice either -- so A is Necessary for them to suffice; moreover, other conjunctions which don't include A, such as P&Q, say, also suffice for B -- so the A-M-N set is Unecessary. Symbolically $(x)(\{Ax \& Mx \& Nx \text{ OR } Px \& Qx\} \rightarrow Bx)$. (Note that we now have a third sense of "necessary" -- A is necessary-for-A&M&N-to-be-sufficient.)

Singular Causation

Hume's Second (Counterfactual) Definition

What is it for a particular instance of A to cause a particular instance of B? (It's clearly not enough just that A is an INUS condit of B, and A and B occur, since the other factors required for A to suffice for B may be absent.)

Hume's 2nd def: "If the first object had not been, the second never had existed".

This is a counterfactual conditional (if not-p, then not-q, where p actually occurred).

Can't we still explain this counterfactual, and hence causation, i.t.o. INUS conditions? I.E: if the actual circumstances contain an US condition for B, and A is a IN part of that US condition. Then the actual circumstances suffice for B, but wouldn't if A were absent.

But there's a problem: the idea is—remove A from the actual circs, hold everything else fixed, and see what follows by law. However, why doesn't this allow, "If there hadn't been the short-circuit, there wouldn't have been the frayed insulation" (because laws tell us that in the circs frayed insulation implies short circuit)? We are in danger of concluding that effects cause their causes. And similarly, "If the barometer hadn't fallen, there wouldn't have been a storm" (because laws tell us that in the circs no barometer fall implies no pressure fall implies no storm). We are in danger of concluding that mere symptoms cause effects.

(NB it's counterfactual/subjective conditionals, not indicative conditionals, that are at issue here: compare "If Oswald didn't kill Kennedy, somebody else did" (certainly TRUE) with "If Oswald hadn't killed Kennedy, somebody else would have" (FALSE—assuming the Warren Commission is right and there was no conspiracy). Similarly "If there hadn't been the short-circuit, there wouldn't have been the frayed insulation" is FALSE, even though in the same circumstances "If there wasn't a short-circuit, the insulation didn't fray" could be TRUE. If we want to analyse causation i.t.o conditionals, we want "If no short-circuit, then no frayed insulation" to be FALSE, and so must focus on counterfactual conditionals, not indicative ones.)

Lewis's Account of Counterfactuals and Causation

Lewis thinks the difficulty of getting counterfactual conditionals right is fatal to a regularity theory of causation. He wants to stick with the connection between singular causes and counterfactuals, but

analyses the latter directly, and not in terms of lawful regularities.

He explains counterfactuals in terms of possible worlds: "If A had been, then B" is true iff (roughly) the nearest A-world is also a B-world.

Lewis also points out that the connection between counterfactuals and causation is a bit more complex than Hume's second definition suggests. Hume said in effect that particular A causes particular B iff If not-A, then not-B. But, while the rhs implies the lhs, we can have cases where A causes B, yet B would still have occurred without A -- namely where A preempts an alternative cause C which would have caused B if A hadn't. (Imagine that overheating causes a valve to open A and thereby stops the pressure increase B. But if the valve had failed to open then the power would have cut out C and caused B anyway.)

Lewis explains why A still causes B here, even though B doesn't counterfactually depend on A, by saying that a chain of events between A and B, each of which counterfactually depends on its predecessor, suffices for causation of B by A. In our example he would thus postulate a D, the release of steam, say, which wouldn't have been there if the valve hadn't opened, and without which the pressure wouldn't have fallen.

But why is it true that without D the pressure wouldn't have fallen? Why can't we argue that without D, the release of steam, there wouldn't have been A, the valve opening, and so C, the power switch, would have been triggered, and B, the pressure would have stopped, anyway. Lewis blocks this by saying that it's false that without D, the release of steam, there wouldn't have been A, the valve opening. But why so?

More generally, why does Lewis assume that effects depend counterfactually on their (direct) causes, but that causes don't depend counterfactually on their direct effects? If the nearest not-C world is a not-E world (If not-C, then not-E), why isn't the nearest not-E world a not-C world? This was the point that made him reject regularity theories, but it is not clear he is any better.

Lewis's Asymmetry of Overdetermination

Lewis recognizes the problem and has an interesting answer.

Consider first this case. "If Nixon had pushed the button (P), there would have been nuclear war (N)". True. But a P and not-N world is surely much closer to actuality than a P and N world (it doesn't have all that mess). Lewis suggests not. For a P-world will have lots of other effects apart from N, and they will have lots of effects, . . . So even a P and not-N world will be very far from actuality, and in addition will require a little miracle, to stop N. So, all in all, it will be further than a P and N world.

You might regard this as ad hoc. But it does lock onto something real, which does distinguish causes from effects non-question-beggingly. The basis of the asymmetry here is a de facto feature of the world,

namely, the asymmetry of overdetermination: overdetermination of effects by causes is very rare, but massive overdetermination of causes by effects is absolutely normal. This means that it's relatively "easy" to "remove" an effect by "removing" its cause—there's nothing else left to fix the effect—but "difficult" to "remove" a cause just by "removing" one effect—since all the other effects will still be there to fix the cause.

This is a good explanation of causal asymmetry, but it can be detached from Lewis's possible worlds account of counterfactuals. Lewis's argument against the regularity account of single-case causation was in effect that it had no grip on causal direction (that's why it has problems with epiphenomena and pre-emption). But he has now offered a way in which regularity theorists can get a grip on causal direction. This means that they can explain counterfactuals and hence singular causation, by saying "Remove C from the actual world, but hold fixed everything that's causally prior to or independent of C, and then see what follows by law . . ."

Why couldn't regularity theorists just use time here? Hold fixed things before and including time of A. But this is unattractive. We don't want to take it for granted that the causal arrow lines up with the temporal one, and so rule out "backwards causation" (time travel, precognition) a priori.

The Direction of Causation

Hume's Temporal Analysis

Problems: (i) can't there be simultaneous causes and effects [inconclusive]; (ii) isn't backwards causation conceivable? [inconclusive] (iii) mightn't we want to explain the direction of time in terms of the direction of causation?

Different Arrows in Time

Assume that time is a dimension. The idea that there is an arrow of time, which goes from earlier events to later ones, is a further idea (NB there are no spatial arrows). The arrow of causation is another arrow which can be imposed on this dimension.

Some philosophers want to explain the earlier-later arrow in terms of the movement of the present from past to future. But this idea seems incoherent (McTaggart, Mellor Real Time). The past-future difference isn't an objectively moving point, but an indexical contrast available from every point in time. But if time doesn't move, we need another way to explain the contrast.

Why not in terms of the causal arrow? Consider: could there be a world in which causes always come later than their effects? But what would it be like to live in such a world? We'd remember the "future" and make plans to affect the "past". But that would be just like the actual world. Turning round the causal

arrow will turn round the temporal arrow.

But now we'd like another arrow to explain causal direction (given that analyses of causation have trouble making it asymmetric).

Lewis's Arrow of Overdetermination

Take any event. In one direction in time there will be lots of events of the kinds it is generally associated with; in the other direction there will be only one such event. This fixes another arrow. Lewis explains the causal arrow in terms of this arrow. he thinks that if this arrow turned around in time, then so would the causal arrow (and the temporal one).

Another way to explain the direction of time is in terms of probabilistic asymmetry. First, by way of introduction to this, let me say something about

Probabilistic Causation

We'd like to allow causes that aren't constantly conjoined with their effect (where not-E wouldn't have been certain given not-C). Hempel suggested that it be enough that C make E highly probable. But smoking doesn't make cancer highly probable. Better (Salmon): C should make E more probable than not-C does: $\text{Prob}(E/C) > \text{Prob}(E/-C)$. Taken as a generalization, this is just the requirement that C and E be correlated, which allows "spurious" causation. So need to add that there be no common cause D which screens off the correlation: no D such that $\text{Prob}(E/C\&D)=\text{Prob}(E/D)$ and $\text{Prob}(E/C\&-D)=\text{Prob}(E/-D)$. Question: can these probabilities merely be reflections of our ignorance, or do they need to reflect to genuine indeterminism?

A Probabilistic Arrow

Lewis's arrow can be put in probabilistic terms (whether or not the probabilities reflect our ignorance, or genuine indeterminism). The different effects of a joint cause are correlated: $\text{Prob}(A\&B) > \text{Prob}(A)\text{Prob}(B)$, but the different causes of a joint effect aren't.

It's a bit tricky to translate this into an explicit analysis of causal direction. But Dan Hausman has an elegant formulation. Take two correlated events A and B. Both A and B will be correlated with lots of other events. If all the events correlated with A are also correlated with B then A must cause B. Conversely, if some events correlated with B are not correlated with A, then again A must cause B.

NB this only works nicely (the last two sentences don't give conflicting answers) if, for every cause-effect pair, there is another cause of the effect that is uncorrelated with the first cause. Question: couldn't there be a world in which there was only one cause and one effect? Hausman: no.

The Relata of Causation

Which kinds of entities are related by single-case causation: events, property-instantiations, facts, or what?

Events

Davidson takes causation to relate events, construed as particulars. Such events can be picked out by many different properties. So the following can all report the same causal truth, for Davidson: The hurricane caused mass destruction; the event described on page 3 of the Times caused mass destruction; the most frightening thing I've ever seen caused mass destruction.

Davidson thinks causation requires laws. But there need only be some description under which the cause-event and effect-event are related by a law; given that, the events can be picked out by other descriptions.

Facts

For Mellor the relata of causation are facts. So for Mellor the basic causal truth is: Much was destroyed because there was a hurricane. By contrast, it's not true that: Much was destroyed because an event was described on page 3 of the Times. The facts in question need to be counterfactually dependent/related by law.

A complication. Kim says that causation relates "events", but that (contra Davidson) events are instantiations of properties. This is actually a version of Mellor's view, not Davidson's. For a particular possessing a property is one kind of fact. But -- for Mellor, and against Kim -- these are not the only facts, and other kinds (existential facts, conjunctive facts, . . .) seem able to enter into causal relationships. (NB. The fact-theory does not require a non-Humean view of laws. There is a connection. Both require properties. But you can have properties, and still be a Humean about laws.)

Causal Explanation versus Causation

Davidsonians need to explain why "The hurricane caused mass destruction" seems "more causal" than "The event described on page 3 of the Times caused mass destruction". They say the former, unlike the latter, is a causal explanation, in that it presents the events via descriptions that enter into laws.

Mellorians need to explain why "The event described on page 3 of the Times caused mass destruction" is true. They say that "Event c caused event e" follows from "An event of kind E occurred because an event of kind C occurred". And they agree that in this construction we can refer to the events by other descriptions. But these particular-event causal truths are derivative from the causal relationships between facts.

The Slingshot

So far a stand-off. But the issue matters to other philosophical topics (esp in phil of mind). You might prefer Mellor because it keeps causation closer to laws/dependence. But Davidson has the notorious "slingshot argument" as a reductio of the idea that "E because C" names facts on either side of the "because" relation. If it did, he says, then we ought to be able to substitute salva veritate the logically equivalent sentence " $\{x: x=x \ \& \ C\} = \{x: x=x\}$ " for "C", thus getting "E because $\{x: x=x \ \& \ C\} = \{x: x=x\}$ ". And then, given any sentence "T" equivalent in truth value to "C" we ought to be able to substitute salva veritate the co-referring " $\{x: x=x \ \& \ T\}$ " for " $\{x: x=x \ \& \ C\}$ ", thus getting "E because $\{x: x=x \ \& \ T\} = \{x: x=x\}$ ". And then, substituting logical equivalents again, we should be able to get "E because T". But it's absurd that "E because T" should be true (or false) whenever "E because C" is and "T" and "C" have the same truth value.

If you want to block this, you need to deny either that the logical equivalent sentences, or co-referring descriptions of sets, can be substituted salva veritate in "E because C".