Causation First?

On the face of the difficulties of counterfactual accounts of causation, some authors have started to explore the idea that we should reverse the order of explanation, and think of *causation as prior to modality*.

Plan for the day:
1. Edgington's arguments for the priority of causation over counterfactual dependence;
2. Some steps of Kment's account of modality in terms of causation.

1. Edgington: Causation is Prior to Counterfactuals

§1. A general statement of the claim that causation is prior, and general reasons to find it plausible.

Causation is implicated in almost everything we think, say and do.

- We ordinarily refer to causation:
  - E.g. many transitive verbs ('push', 'break', 'knock over', 'hurt'…)
  - E.g. many adjectives ('soft', 'hard'…); E assumes that something is hard to the extent that it cannot be easily made to change.
- We perceive causation (this takes away one of the motivation for trying to reduce it to something else):
  - E.g. cutting, jumping, running: perceiving that is perceiving the efficacy of causation.
  - The perception is of course fallible: there can be visual illusions of causation.
    - Suppose I'm watching a fork-lift truck raising a metal box. As it seems—I am then told that the box is actually being pulled up by a magnet, the truck is exerting no force (it is merely a back-up device). The illusion still persists: it looks just as if the truck is lifting the box. I do not think there could be knowledge or experience of a world which is not causally structured.

Being causally structured is a necessary condition of a world in which there can be knowledge, experience, and action. So it is not surprising if causation is so basic that there will not be any illuminating analysis of the form: c causes e iff ….

§2. Counterfactual accounts of causation as problematic as regularity accounts of causation.

The attempt to define causation in terms of counterfactual conditionals (roughly, *c* causes *e* iff, if *c* had not occurred, *e* would not have occurred) is beset with very much the same difficulties as the attempt to define causation in terms of regularities or laws:

A. The problem of distinguishing causes from effects.

   Edgington discusses a passage from Lewis (1973). If the pressure had not been *p*, the barometer would not have read *r* (so, the pressure's being *p* caused the barometer's reading *r*). What about the reverse? This is what Lewis says about this example:

> **BAROMETER**. If the reading had been higher, would the pressure have been higher? Or would the barometer have been malfunctioning? The second sounds better: a higher reading would have been an incorrect reading. … When [we suppose a higher reading], it is less of a departure from actuality to hold the pressure fixed and sacrifice the accuracy of the barometer, rather than vice versa. It is not hard to see why. The barometer, being more localized and more delicate than the weather, is more vulnerable to slight departures from actuality.

The two counterfactuals Edgington has in mind are:

> (1) If the reading had been higher [or: had not been *r*], the pressure would have been higher.
> (2) If the reading had been higher [or: had not been *r*], the barometer would have been malfunctioning.

Lewis's claim: in the context he describes, (1) is false and (2) is true.

Edgington's points:

a) In many contexts (1) is true (and this remains an issue for Lewis) – imagine a very sturdy barometer;

b) It's not obvious that (2) is true in the given context (the weather is fragile, after all);

c) if (2) is true as Lewis says, then his counterfactual theory of causation applies; *prima facie*, here it entails that the reading *r* of the barometer causes the barometer to function properly. But this is not right. *[Since Victor and Zepo helpfully asked about (c), the point is here explained in an expanded version – it was shorter in the handout you got in class.]*

B. Epiphenomena.

(3) If Fred had got ill, we would have got ill too.

> **PIE**. We all ate the same pie for dinner. Fred got ill and died. The question is whether the cause of death was food-poisoning from the pie. No it, is argued: suppose the pie was poisonous. Then if Fred had got ill, we would have got ill too (or at least that is quite likely), and we didn't. We test a drug on rats and conclude that it is safe for humans, because, if it were unsafe for humans, it would be unsafe for rats.

The counterfactuals are quite in order, yet they have the same structure as arguing from spots to fever, or from barometer to rain.

C. Pre-emption. (Covered last week.)

§3. Take the class of 'standard' counterfactuals concerning matters of particular fact, the consequent being later in time than the antecedent—the class that is most promising for an account of causation. In assessing it, we presuppose causation.

- Specifically, E. claims, we cannot give an account of what we hold constant in assessing these counterfactuals, without *appealing* to causal notions such as causal dependence or independence.

  For instance, consider:

  (4) If I had bet on heads I would have won.

  > **COIN**. I decline to bet on the toss of a coin. It is tossed anyway. It lands heads. *If it had bet on heads I would have won*. This is so provided that the toss and my betting are causally independent of each other: say the toss takes place in one room, I write Heads or Tails or No Bet on a piece of paper in another room and there is no causal interaction between these events.

  In assessing (4), we keep constant the fact that the coin did land heads, provided that the toss is causally independent of the bet. But what if we don't assume causal independence?

  > E: "If on the other hand, my betting might have caused you to toss the coin a little later, or a little differently, the counterfactual doesn't hold."

  Similarly:

  (5) If I had caught the plane I would be dead.

  (5) is true *provided* the causal story of the crash is independent of my absence from the plane. Here is a context on which (5) is true:

  > **PLANE.** The car breaks down on my way to the airport. I miss the plane. Later I discover it crashed: *if I had caught the plane I would be dead*. Suppose that a chance event, not predictable in advance, brought down the plane. At the time of take-off, the plane was not relevantly different, with respect to safety, from any normal plane: there was a small but non-zero chance that some such accident would occur—due to freak weather conditions, or freak electrical or mechanical faults (or combinations thereof), or a freak heart attack or heart attacks on the part of those in control. Suppose also that my absence from the plane had no effect on the causal history of the crash: it's not the case that, e.g., some subtle feature of the distribution of weight contributed to the crash, which might have been different, had I been aboard.

  Can you think of a context on which the crash is dependent of my absence from the plane, so to make (5) false?

  In short: according to E., we keep constant later features of the actual world, provided that they are causally independent of the antecedent.

Interesting point made in class by Beatrice and Caspar: Lewis seems to be able to get the truth-value of the counterfactual right for COIN and PLANE. One may think that our

using causal reasoning in assessing them is not a reason to conclude that causation needs to feature in the semantics of counterfactuals. If so, §3 of the paper only gives us prima facie *motivation for thinking that causation is prior to counterfactuals. (But what about the cases of §2, especially BAROMETER?)*

§ 4. Not all counterfactuals track causation. Counterfactuals are too wide a class to hope to capture causation in terms of them.

For one thing, Edgington points out, 'If A happens, B will happen but A won't cause B to happen' is never contradictory – but some conditionals are rightly presumed to be asserted on causal grounds. So we have a semantic mismatch.

Moreover, there are causes of counterfactual dependence that are not cases of causation. Consider:

(6) If it hadn't been in the attic it would have been in the garden.

> **TREASURE**. There is a treasure hunt. The organizer says: 'I'll give you a hint: it's either in the attic or the garden'. Trusting the speaker, I think 'If it's not in the attic it's in the garden'. We are competing in pairs: I go to the attic and tip off my partner to search the garden. I discover the treasure. 'Why did you tell me to go the garden?' she asks. 'Because if it hadn't been in the attic it would have been in the garden: that's (what I inferred from) what I was told'. That doesn't sound wrong in the context.

Given TREASURE, (6) is true. But this counterfactual dependence doesn't imply causation.

Questions to think about:

- Should we agree with Edgington? Why?

- What can Hume and Lewis answer to her respective criticisms, if anything?

2. Kment: Modality in Terms of Causation/Explanation

Kment goes a step further than Edgington: he inverts the order of explanation and accounts for modality in terms of causation – or rather, ontic explanation.

Kment's first steps rely on Edgingtonian examples:

> **LOTTERY 1**. As you are about to watch an indeterministic lottery on tv, someone offers to seel you ticket number 17. You decline. The ticket wins. It seems true to say that 'if you had bought the ticket, you would have won'.
>
> **LOTTERY 2**. The company organizing the lottery described in LOTTERY 1 has two qualitatively indistinguishable lottery machines giving the same chance to every possible outcome. They used machine A in the draw. But could have used B. It seems false to say that 'if a different machine would have been used, 17 would still have won'.

Kment's interpretation of Edgington's point about presupposing causal dependence:

> "[T]he use of a particular lottery machine is part of the causal history of the outcome. Hence, which machine is used makes a difference to the causal history of the result. That's why the outcome of the draw cannot be held fixed in the second case. (MER, p. 7).
>
> So counterfactual dependence seems to go with causal dependence here.

- **<u>How to import this insight in the semantics of counterfactuals?</u>**
  K. keeps a Lewisian framework of closeness:

  > 'A > C' is true iff some A-world at which C is true is closer to actuality than any A-world at which C is not true.

  The ordering of worlds by their closeness to actuality is also once again the result of weighing their respective similarities to actuality against each other.

  So where is the account different from Lewis's?
  Kment differs from Lewis in his account of what similarities matter to closeness.
  For Kment, this depends at first pass on causal relevance:

  <u>The causal criterion of relevance</u>

  > (CCR) Similarities between two worlds matter to the closeness ordering just in case they concern facts that have the same causes at the two worlds.

Kment thinks that this criterion generalizes to a broader notion of ontic explanation. Consider:

(7) If the law of gravitation hadn't been a law, events would still have conformed to it.

For Kment, (7) is false: events conform to gravitation in actuality because of its being a law, where 'because of' tracks metaphysical (or ontic) explanation.

The explanatory criterion of relevance (first pass)

(ECR – first pass) Similarities between two worlds matter to the closeness ordering just in case they concern facts that have the same explanations at the two worlds.

Examples of explanation

The absence of safeguards caused the short circuit to occur

The fact that it's a law that all Fs are G explains the fact that all Fs are G.

The fact that Fred is an atom with atomic number 79 grounds the fact that Fred is a gold atom.

Covering-law conception of grounding

Grounds are connected to the facts they ground by metaphysical laws.

The metaphysical laws include essential truths. For example, it's an essential truth about the property of being a gold atom that

(8) For all x, x is a gold atom iff x is an atom with atomic number 79.

The explanatory criterion of relevance (second pass):

(ECR - revised) If some fact *f* obtains both at the actual world and at world *w*, then this similarity is relevant to the closeness ordering iff every fact *g* that forms part of *f*'s explanatory history obtains at *w*.

From this, Kment develops the following:

Standards of closeness in Kment's framework:

1. It is of the first importance to maximize match with respect to metaphysical laws.
2. It is of the second importance to maximize match in the natural laws.
3. It is of the third importance to avoid large alien violations of the actual laws of nature.
4. It is of the fourth importance to maximize match in matters of particular fact.
5. It is of the fifth importance to avoid small violations of the actual laws of nature.
6. It is of the sixth importance to avoid departures. *[Cf. MER, ch.12]*

Let's look at it with an example:

(9) If Fred had filed his tax return two hours before the deadline, he would not have been penalized.

How do we assess this given K's criteria?

An Important Caveat: Kments Laws Tolerate Exceptions

Consider:

> (10) If I had not scratched my nose a minute ago, my nose wouldn't be bothering me now.

Problem: Assume determinism is true, and that the state of the universe at any moment before scratching together with the laws determines that I scratch my nose. Then any metaphysically possible world where I don't scratch my nose is either unlike actuality throughout the pre-antecedent time, or features some violations of the laws (or both).

This means that one of these two counterfactuals has to be true:

> (11) If I had not scratched my nose a minute ago, the history of the world might have been different throughout the history before the scratching.

> (12) If I had not scratched my nose a minute ago, the laws of nature might have been different.

But they both seem false.

Kment's solution: allow that the laws may have exceptions.

That is, Kment rejects the *factivity of laws*:

> LAW (All Fs are Gs) → all Fs are Gs

The closest world in which I don't scratch my nose has the same laws, but a small violation of these laws occurs there, because of which I don't get to scratch my nose.

Two possible readings of Kment's claim that in the relevant scenario "the laws are the same" (this came up in class). The expression 'the laws are the same' is ambiguous between:

i) *L* is a law at a world which can be violated at that world; it is a false at that world when violated at that world.
ii) *L* is a law which can be violated at a world, but remains true at that world even when violated at that world.

We weren't sure about Boris's preference between i-ii.

Question to think about: Is denying the factivity of laws the best solution for Kment? Why? Or could he somehow plausibly endorse a picture on which the ECR holds and laws are factive?

- *A good outcome of Kment's picture:*

The semantics seems able to accommodate cases like Elga's counterexample to Lewis (discussed last time).

- *Some remaining concerns about this picture:*

a. Some counterexamples to Lewis's account of counterfactuals still seem to apply.

(13) If the coat had been stolen last night, it would have been stolen at midnight.

> **MIDNIGHT** (from Pollock). I forgot my coat at the bar last night. In the course of the night two potential coat thieves passed the coat, one at 10pm, the other at midnight. Each time, there was a non-zero chance that the coat would be stolen. The next morning, I find to my relief that the coat is still where I left.

Kment's semantics, as he admits (MER, p. 242) leads one to conclude that the closest world the coat is stolen at midnight, so that (13) comes up true.

b. Some worry that Kment's semantics also open itself up to objections Lewis's didn't suffer from:

(14) If Paula had used phone B (rather than A) to make a phone call, the outcome of the lottery draw would have been the same.

> **INDETERMINISTIC LOTTERY.** You are watching an indeterministic lottery draw. The lottery was instituted ten years ago. As part of this process, one of the people in charge (Paula) made an important phone call. She has two qualitatively identical phones, A and B, and she always chooses one of them at random when she has to make a call. When she happened to make the aforementioned phone call 10 years ago, she happened to use A.

Given this scenario, (14) seems true. But the actual explanatory history of the outcome of the draw includes certain facts about A that don't obtain at the closest antecedent-worlds. So, by ECR, similarities in the outcome of the draw aren't closeness-relevant. But then, it is not guaranteed that (14) comes out true on Kment's semantics.

Kment's response to this last worry: minimizing closeness-relevant dissimilarities (or departures). (cf. MER, p. 234.)

Question to think about

- Are we happy with Kment's answer to b? Why?
- Is there an answer we can give to (a) on Kment's behalf? And are we happy with his answer to (b), or is it too *ad hoc*?