Lecture 4
The Asymmetry of Causation

1. Introduction

One puzzle in the metaphysics of causation concerns its **time-asymmetry**. That is, its tendency to 'move' from past to future. The tendency for causes to be in the past relative to their effects.

Some, such as Reichenbach and Suppes (and classically, Hume), make this characteristic part of their analyses of causation. What it is to be cause, on these views, is (at least in part) to occur prior to its effect.

However, others argue that this is unsatisfactory. This strategy makes it a conceptual truth that causes occur before their effects. But there does not seem to be any contradiction in supposing that some cause occurs *after* its effect, or even *simultaneous to* its effect. Those who share this view attempt to give non-causal and non-temporal explanations for the time-asymmetry of causation.

Some of these strategies involve appealing to a '**third arrow**' that is a common explanation for both time's arrow (that is, time's asymmetry) and causation's arrow (i.e. causation's time-asymmetry). We'll look at a few different candidate third arrows.

2. The Arrow of Counterfactual Dependence

Lewis argues that the time-asymmetry of causation arises from the asymmetry of counterfactual dependence. And this latter asymmetry is *contingent*. It so happens that, at our world, future events tend to counterfactually depend on past events, and not vice versa.

Here again are Lewis' conditions for ranking the similarity of worlds:

(I) Avoid big, widespread, diverse violations of law.
(II) Maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
(III) Avoid even small, localised, simple violations of law.
(IV) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

According to Lewis, we would need many many small, localised simple violations of law in order to secure a match of facts to the actual world *after* a counterfactual event than we would to secure a match of facts to the actual world *before* a counterfactual event. That is, our world is one where, once an event happens there are lots of "*determinants*"

of that event. For instance, to use the egg-breaking example again; consider what the world will contain in the future with respect to the event of breaking an egg:

- the change in the particles in the air caused by the vibration of the egg striking the side of the pan
- the memory I have of the egg's breaking.
- the various bits of egg shell that are projected in particular directions because of the exact way the egg struck the pan
- the change in arrangement of the yolk and white in the shell because of the strike
- the slight movement of the pan caused by the force of the egg on its edge

But the same is not true of the world in the past with respect to that event. There are far fewer 'traces'.

The argument is, a world that is just like ours up until a point immediately before the breaking of the egg and then **diverges** will only require one small violation of law (immediately before the point when, at the actual world, I broke an egg) for it to be the case that I didn't break the egg.

On the other hand, a world that is like ours up until a point immediately before the breaking of the egg and the **converges** will require many small violations of law to 'implant' the contents of our world like those listed above.

Given that worlds that converge in the past and diverge in the future relative to our world are closer than worlds for which the converse is true, it will turn out that counterfactuals of the form 'if *x* had occurred, *y* would have occurred in the future' will be true but those of the form 'if *x* had occurred, then *y* would have been the case in the past' will be false. This is the asymmetry of counterfactual dependence. And since, on Lewis' account of causation causal relations are chains of counterfactual dependence, it is this contingent asymmetry that explains why causation at the actual world is time-asymmetric.

Objection

Adam Elga (2001) argues that when one considers macroscopic thermodynamic processes carefully enough, it reveals that Lewis is wrong about the asymmetry of counterfactual dependence.

Lewis argued that, relative to our world, it takes more small violations of law to secure convergence of events into the future at another world than it does to secure convergence of events into the past at that world. Elga shows that, given the laws of

thermodynamics, it is possible to secure convergence into the future without positing many violations of law.

Elga presents the following case. Consider Gretta, at the actual world, who is about to crack an egg into a hot frying pan. The microphysical state of the world at S_0, in which Gretta is about to crack the egg, evolves into S_1, in which the egg is cooked and sitting on the pan. From this we can define the following process:

> **The S_0-to-S_1 process** The egg oozes out of the cracked shell and drops down towards the pan, where it splats on the pan, making a noise, slightly heating up the surrounding air, and setting up some vibrations in the pan. Then the egg cooks by absorbing heat from the pan.

Elga then defines Z_1, what he calls the "*velocity-reverse* of S_1" (318, original emphasis). In Z_1, all of the particles are just like those in S_1 except each one's velocity is reversed such that, when you evolve Z_1 into the future it evolves in the way that S_1 would if you evolved it backwards. This results in the state Z_0, which is just like S_0 except the velocities of the particles are reversed. From this we can define the process:

> **The Z_1-to- Z_0 process** The cooked egg uncooks by giving up heat to the (already very hot) pan. Meanwhile, molecules in the pan start to coordinate to form a pattern of vibration converging on the center of the pan. Air molecules around the room begin to form a series of spherical waves that converge on the pan. Just as the egg finishes uncooking, the coordinated action of these inward-directed air waves and pan vibrations congeals the egg into a round shape and propels it vertically towards the waiting open shell, which then seals around it.

The laws that govern these two processes are exactly the same. Just imagine that the first is something like playing a tape forwards, and the second like playing that same tape backwards. The relationships between the particles and their movements are the same, they just occur in the reverse order. **Both scenarios respect our dynamical laws**.

Now, the states named above are *microphysical* descriptions of the world. They give descriptions of the precise location of each particle. We can also talk about macroscopic events. Elga defines 'COOKED' as the "set of states that are exactly like Z_1 with respect to coarse-grained macroscopic parameters (such as temperature and pressure distribution)" (319). COOKED is what we call ***multiply realisable***. There are multiple different microscopic states that can realise the macroscopic state.

Some of the states in COOKED have normal futures. They have futures like we'd expect them to have—e.g. wherein the egg sits in the pan and cools over time. Other states in

COOKED, though have **abnormal futures**. They have futures like Z_0, wherein the egg uncooks, comes together, and leaps back into the shell. There are many more normal futures than abnormal ones.

Elga argues that the Z_1-Z_0 process is very "fragile". That is to say, any small change to Z_1 would have the result that it likely didn't evolve into Z_0. By analogy, think about it in terms of paths between two locations. Suppose you're trying to get from A to B. If there is only one very narrow path from A to B, and your starting position at A is changed ever-so-slightly, you'll miss the path and fail to reach B. On the other hand, if there is an absolutely massive path from A to B, a small change in your starting position won't bump you off the path. It would be much harder to bump you off such a path.

But if this is so of Z_1, then it is also true of S_1, since S_1 was just like Z_1 except for the directions of the particles. And a small 'bump' to the state Z_1 resulting in a future where the egg does not leap into the egg corresponds to a small 'bump' to the state of S_1 resulting in a *past* where the egg is never cracked. In other words, we could insert a small miracle at a world that is different from the actual one until a time just after when Gretta actually cracked an egg, and then that world could converge with the actual world. No large (or set of many) violations of law required.

But if there is no asymmetry in the nature of these violations of law, and that asymmetry is what guaranteed the (local) asymmetry of counterfactual dependence, then there is no asymmetry of counterfactual dependence. And if there is no such asymmetry of counterfactual dependence, then this cannot explain the asymmetry of causation.

3. <u>The Arrow of Entropy (A Fourth Arrow?)</u>

This strategy involves explaining why time is asymmetric and why causation is time-asymmetric by appealing to the direction of entropy. The argument goes, even though our microphysical laws are time-symmetric, the thermodynamic laws are not. Thermodynamic laws then provide a viable means of explaining each of the other asymmetries in question. In explaining them independently, the fact that the two share a common explanation would account for the fact that they both 'travel' or 'point' in the same direction.

For those who also attended the Time lectures, the account here will be familiar. The combination of the thermodynamic laws **and** the Past Hypothesis (the hypothesis that there is a point of very low entropy in the very distant past) combine to yield the result that macro-physical processes evolve from past to future.

Thus, causation is time-asymmetric because, at the macroscopic level, the Past Hypothesis and the Laws ensure that "small, local changes […] produce much bigger and

more diverse changes in the future than they do in the past" (Price and Weslake 2010: 424). This is just what we needed for the asymmetry of counterfactual dependence! From here, insert Lewis' story from above, and the explanation of causation's time-asymmetry is complete.

It is worth noting that you could also apply this strategy to a probabilistic account of causation (and not just to a counterfactual one). Field (2003) argues for just this. Field argues that there is an asymmetry in probabilistic dependence—specifically in the dependence relation that respects the relations modeled in causal graphs (like those we saw from Woodward (2003)).

Objection

One objection to this view is that positing a Past Hypothesis does not guarantee that all causes will be time-asymmetric in the way that we expect them to be. That is, it is consistent with there being a moment of very low entropy in the past that entropy decrease at local areas in spacetime. But, if entropy occasionally decreases, and it is the *increase* in entropy that is meant to explain the asymmetry of time, and the time-asymmetry of causation, the it follows that time and causation 'move' in the opposite when entropy decreases, and reverse again when entropy increases.

So, there is a **metaphysical** objection concerning the nature of causation on this view. This view allows for the possibility that, at some points in spacetime, all (or most) causal relations operate in the opposite temporal direction, and at other points in spacetime, causation reverses direction.

Now, you might not think this is a problem. There is nothing in the concept of causation that makes this inconsistent. It is unfamiliar to be certain, but that does not entail that it is impossible.

However, there is another kind of objection raised on this basis: we can call it a **subjectivist** objection. Here is the worry from Price and Weslake: our concept of causation is intimately related to our practices of deliberation. Our practice of deliberation is such that we consider performing actions in order to bring about outcomes in the *future* relative to those actions. We never act in order to change the past. As such, whatever explains the time-asymmetry of *causation* should also be able to explain the **time-asymmetry of deliberation**. If this is right, then the entropy-based explanation fails. Price and Weslake argue that if it were the case that either a local or a global increase in entropy towards the future explains the time-asymmetry of causation and of deliberation, it should also be the case that, if entropy increased towards the past, we would deliberate now to affect past events instead.

But it is not the case that future constraints stop us being able to being able to exercise control over the future. Here are Price and Weslake in their own words:

> The first question is whether such a future constraint would imply that our deliberative phenomenology would be a future-directed analogue of what we are trying to explain with respect to the past: the sheer apparent absurdity, at least in ordinary cases, of acting so as to influence the past. It is hard to see why this should be so. **Restrictions in the distant future—even extreme restrictions, much tighter than PH itself—seem to have virtually no bearing on our present sense that we can affect the future**. Suppose God tells us that as a matter of law, the final state, some fifteen billion years from now, will be constrained within some tiny region of phase space (comparable in size to that required by PH). […] Do we lapse into fatalism, coming to think it absurd that we might seek to influence our immediate future? It is hard to see why we would, or should. Hence, by symmetry, it is hard to see why a remote past hypothesis should be incompatible with taking ourselves to be able to affect the near past. (425, my emphasis)

To illustrate this on a smaller scale, the authors present a variation on the Death in Damascus case from decision theory.

> […] suppose we believe that we are destined to meet Death at noon on a certain day. […] It is now 09:05 on the fateful morning, and we sit in Aleppo airport, with a boarding pass for the flight to Damascus. We know that Death will meet us in one place or other; and moreover (since he refuses to fly) that he is already on the road to whichever place it is to be. Is it absurd to think that we are still free to choose whether to board the plane? On the contrary, apparently. While the boundary condition certainly deprives us of many options—the option to be anywhere other than Damascus or Aleppo at noon, for example, or to be anywhere at all, later in the day—**it also yields some new abilities**: **in particular, the ability to influence Death's movements, even somewhat earlier on the day in question**. (425-6, my emphasis)

From this they argue that, when there is a future constraint (like our destined appointment with Death), we can affect the future in limited respects, and this can even result in our affecting the *past* in some respects.

These two points suggest the following, according to Price and Weslake:

- A point of low-entropy in the future—a future constraint—would not have the result of inclining us to fatalism. Whereas, we think that it is *absurd* to consider acting to change the past. This suggests that it is not the Past Hypothesis that is generating this sense of absurdity, since an analogous condition in the future produces no such sense.

4. The Arrow of Deliberation

Price and Weslake argue that we should understand the asymmetry of causation *in terms of* the asymmetry of deliberation. Indeed, they think that this is the only way to guarantee that there is a good explanation for why we always deliberate from past to future, and in a way that 'lines up' with the time-asymmetry of causation.

We are "agents for whom actions *follow* deliberation" (434). From this perspective, the future of events on which we deliberate is asymmetric to the past of those events, since in the past our deliberation lies, and in the future it does not. And, given that this is the perspective from which we reason about causes, this is also what explains the asymmetry of causation. If A is a cause of B just in case an agent's credence in B given that they bring about A is great than their unconditional credence in B, and if considering A to be a potential action (i.e. *deliberating* about A) involves imposing our perspective (i.e that perspective such that deliberations are in the past relative to actions), then causes will be considered to be prior to their effects from our perspective.

This explanation is meant to have the advantages that it does not make causation time-asymmetric by fiat, it explains why a number of different asymmetries 'line up' in the same direction, and it doesn't face the objections that faced the previous two strategies.