METAPHYSICS AND EPISTEMOLOGY

Causation

1. We shall deal with the causal relation between events. When something x brings about something y, x and y are very often events. E.g. if a stone falls on a glass, and the glass breaks up, we may say that the stone made the glass break; but it is more accurate to say that what made the glass break, the cause, wasn't the stone itself, but its fall on the glass — that fall is an event. We shall mainly deal with efforts to analyse (that is, define) the concept of causation.

Many events are composite; that is, they are made up of other, shorter events. E.g. a war is made up of battles, army movements, etc. (The word 'event' is more appropriate for short incidents than for something that lasts as long as a war, but we will ignore that in our terminology.)

We can distinguish between events themselves, which are particulars (whether they are simple or composite), and event-types. Examples of such types: death, war, recuperation, etc. Each (particular) event has a specific position in time and cannot be repeated. Correspondingly, we distinguish between *singular causal statements* and *general causal statements.* A singular causal statement is a sentence (true or false) of the form '**a** brought about [caused, brings about, etc] **b**' where the terms **a** and **b** refer to (particular) events. E.g. 'The first world war caused the collapse of monarchy in Germany'. A general causal statement is a sentence (true or false) of the form '**a** brings about [causes, etc] **b**' where the terms **a** and **b** refer to event-types. E.g. 'War results in destruction', 'Smoking causes cancer'. It is not clear how (or if) general causal statements can be analysed into statements about the causal relation between particular events. So 'War results in destruction' seems to mean 'Each war results in some destruction'. But 'Smoking causes cancer' doesn't mean 'Each event that is a case of smoking causes the appearance of a carcinoma'. We shall here deal with singular causal statements.

2. Modern discussions of causality often start from D. Hume's views (18^{th} century). According to him, when we say that x caused y, very frequently what we mean is that x preceded y temporally and rendered y necessary, inevitable. Hume considers that this concept of causation (that is, causation = temporal sequence + necessitation) corresponds to nothing outside our minds. The view in question is due to his empiricism. For example, we see billiard balls collide and then change direction; we do not see the collision render the change of direction necessary.

In Hume's view, there is also another concept of causation, and this corresponds to something outside our minds. Hume offered two definitions of that other concept. According to the first definition, 'x caused y' means 'y followed x temporally, and everything similar to x is followed by something similar to y' (that is, causation = temporal sequence + constant conjunction). According to the second definition, 'x caused y' means 'if x hadn't existed, y wouldn't have existed'. Hume treated the two definitions as equivalent, but it's not at all obvious that they are equivalent. Certain contemporary views about causation come from the former definition, and others come from the latter. (Here I have somewhat simplified Hume's concepts and definitions.)

3. According to a view, the claim 'Event x caused event y' means this: y followed x

temporally, and the events x and y satisfy some predicates **F** and **G** respectively such that the statement 'Each **F** is followed by a **G**' constitutes a natural law or is a logical consequence of the natural laws. Variants of that view are adopted by J. S. Mill, D. Davidson and others. E.g. if x is the fall of a stone on a glass, and y is the breaking of the glass, then the event x satisfies the predicate 'exertion of such-and-such forces on material with such-and-such properties' (imagine more specific phrases here), the event y satisfies the predicate 'breakage of the material', and the sentence 'Each exertion of such-and-such forces on material with such-and-such properties is followed by breakage of the material' is a logical consequence of the laws of nature. (A thing *satisfies* a predicate iff the predicate is true of the thing. Also, if we omit the 'is' from a one-place predicate beginning with 'is', e.g. 'is breakage of the material', what remains, e.g. 'breakage of the material', is still called *a predicate*.)

That view is based on Hume's definition to the effect that causation = temporal sequence + constant conjunction. Yet it incorporates something from the idea of necessitation. For if the statement 'Each **F** is followed by a **G**' constitutes a natural law or is a logical consequence of the natural laws, then it is physically necessary that each **F** is followed by a **G**. 'It is physically necessary that p' means 'It couldn't have been that the same natural laws governed the universe as in fact govern it but it was not the case that p'.

In each definition, we distinguish between the *definiendum* (e.g. 'event x caused event y') and the *definiens* (e.g. 'y followed x temporally, and …'). If one wants to argue against a definition, one must find either cases (actual or just logically possible) in which the definiendum is true but the definiens is not or cases in which the definiens is true but the definiendum is not. In the following, we shall see some difficulties for the view we are discussing. Of those, in (a)–(c) it seems that the definiendum is true but not the definiens, whereas in (d) it seems that the definiens is true but not the definiendum.

(a) Prima facie it seems that the analysis we are discussing does not apply to the singular causal statements we come across in social sciences and history. The first world war caused the collapse of the monarchy in Germany, but it doesn't seem that the first world war (or the collapse of the monarchy) satisfies any predicate that occurs in a law of nature.

Supporters of the analysis will object that even an event like the first world war satisfies at least one such predicate, but we do not know what that is. They will say that it is probably a very complex predicate that analyses the composite event into the events that make it up (movements, shootings, etc.) and describes each one of them in terms drawn from the natural sciences (neurological terms if it is a human decision or thought).

(b) According to contemporary physics, sometimes an event e_1 causes some event e_2, but it is not the case that each event of the same type as e_1 is followed by an event of the same type as e_2. Events of the former type are usually, but not always, followed by events of the latter type. Physical theory can calculate how usual that is and hence what is the probability of an event of the latter type following once an event of the former has taken place.

Here, supporters of the analysis we are discussing may modify their definition and, instead of talking about a statement of the form 'Each **F** is followed by a **G**', talk about a statement of the form 'Each **F** has an n% probability of being followed by a **G**', also saying something about how close to 100 n must be. Whenever one event brings about another, the two events come under a regularity of nature, but the regularity may be probabilistic and not unexceptional.

But then, the question arises: isn't it logically possible for an event to cause another

but the two events to come under no regularity of nature, not even a probabilistic one? Many people find it logically possible that an event e_1 should cause an event e_2 but there should be no laws of nature (not even probabilistic ones).

(c) It doesn't seem right to consider that by definition the cause temporally precedes the effect. If that is so by definition, it is not logically possible that the cause and the effect should sometimes be simultaneous. Yet this appears to be logically possible. Moreover, it has been argued that it is logically possible for the cause to occur sometimes after the effect; indeed, there are some theories in contemporary physics according to which that is sometimes just what happens.

Here supporters of the analysis may modify their definition and say that x caused y iff the two events satisfy two predicates **F** and **G** respectively such that the statement 'Each **F** is accompanied by a **G**' constitutes a natural law or is a logical consequence of the natural laws. (I use 'is accompanied' in such a way that, by saying that an event was accompanied by another, I do not specify which event preceded, nor do I specify whether the events were successive or simultaneous.) There emerges, however, a new problem. Let's say that y is the entry, at a certain moment, of the yellow fever virus into the organism of a specific person, and that x is the illness that resulted. As yellow fever cannot occur without the virus, the statement 'Each case of yellow fever is accompanied by an entry of the corresponding virus' is a logical consequence of the natural laws. Hence, according to the modified definition, x brought about y. But in fact y caused x and not the other way round. This is the problem of the effect that could not have been caused differently.

(d) T. Reid (a contemporary of Hume's) adduced the following counterexample to the definition that equates causation with temporal sequence + constant conjunction: each day is followed by a night (and each night by a day), but the days do not bring about the days (nor the nights the days).

Of course, days and nights are not events, but we can find similar counterexamples involving events. Let's take a well-functioning barometer. Whenever its pointer goes below a certain limit while there is moisture, there follows rain. The statement 'Each descent of the pointer below limit *l* under conditions of moisture is followed by rain' is a logical consequence of the laws of nature. However, the rain is not caused by the descent of the pointer. This is the problem of epiphenomena: the descent of the pointer is an epiphenomenon (that is, an effect that doesn't have significant effects itself) of the real cause of the rain.

4. Another view is based on the second definition offered by Hume about causality outside our minds. This view, put forward by D. Lewis, tries to explain what it means to say that an event x is a cause of an event y.

There is a difference between the claim 'x is a cause of y' and the claim 'x causes y'. Sometimes, especially in history, an event has many causes, but none of them on its own caused the event; the event was caused by all the causes acting together. E.g. the second world war had many causes; what caused it was the bundle of those causes (all of them together).

Lewis first introduces the concept of *causal dependence.* Let's take two actual and distinct events x and y. By saying 'y depends causally on x', he means 'If x hadn't occurred, y wouldn't have occurred either'. (Lewis believes that there are also events that are not actual, but merely possible. We can confine ourselves to actual events.) In his view, the statement 'The event x is a cause of the event y' means 'There is a series (not necessarily temporal) of events e_1, …, e_k where e_1 is x, e_k is y, and, for each n, event e_{n+1}

depends causally on e_n'. E.g. the drinking of hemlock was a cause of Socrates' death because there is such a series of events (which are to do with the consecutive results of that drinking to his organism).

The sentence 'If x hadn't occurred, y wouldn't have occurred either' is a *counterfactual conditional.* A counterfactual conditional is any sentence of the form 'If **A** then **B**' which tells us what *would be the case,* or what *would have been the case,* if things were, or had been, a certain way. (Lewis also developed a theory about the logic and semantics of those conditionals.)

According to Lewis, the statement 'If x hadn't occurred, y wouldn't have occurred either' implies the statement 'x is a cause of y'. For if the former statement is true, there is a series of events (the series x, y) of the kind that, in his view, suffices for there being causation. But he considers that the latter statement does not imply the former. He believes that, in some cases, an event x is a cause of an event y although if x hadn't occurred, y would have occurred because of incidents which in fact were not causes. E.g. both A and B shoot C, but A shoots earlier, so C falls dead from A's bullet, and that is why B's bullet doesn't find him. A's shooting is a cause of C's death, but even if A's shooting hadn't occurred, C's death would have occurred, because of B's shooting. Such cases are called 'preemption'.

Now, here are some difficulties for Lewis's view:

(a) Let's say that x was a cause of y and y wasn't a cause of x. Let's also consider that it is not physically possible for x to occur without y occurring. (Something is physically possible iff it is allowed by the laws of nature.) For example, x may be the fall of the atomic bomb on Hiroshima, and y the destruction of the city. It seems that, in this case, it is right to say that if y hadn't occurred, x wouldn't have occurred either. Hence, according to Lewis, y was a cause of x. That is wrong, though.

Lewis replies that at least if x preceded y temporally, it is not valid to infer from the premiss 'It is not physically possible for x to occur without y occurring' to the conclusion 'If y hadn't occurred, x wouldn't have occurred either'. He considers that if y hadn't occurred, then up to that moment the universe would have evolved as it evolved in the actual world, and so x would have occurred; but in the end something would have occurred which was not in agreement with the natural laws of the actual world, and so y would have been averted. This reply presupposes certain aspects of Lewis's theory about the semantics of counterfactual conditionals.

(b) J. Kim emphasized that, for many events x and y, it is true that if x hadn't occurred, y wouldn't have occurred either, but it is wrong to say that x was a cause of y. E.g. I write 'Athens'; if I hadn't written the letter 't', I wouldn't have written 'Athens'. I open the door by turning the handle; if I hadn't turned it, I wouldn't have opened the door. If Peter had not been born in 2000, he wouldn't have turned 21 in 2021. If my sister had not given birth, I wouldn't have become an uncle. In none of those examples do we have a relation between cause and effect. (In the examples of Peter and my sister, we have logical implication rather than causation.)

Lewis does not reply explicitly, but I think he would say the following. For the examples of 'Athens' and the door, he would say that the inference from 'If x hadn't occurred, y wouldn't have occurred either' to 'x is a cause of y' is valid for distinct events. He would add that, by talking about distinct events, he means events such that they are not identical, neither is part of the other, and they have no common part. In those examples, the one event (the writing of 't', the turning of the handle) is part of the other (the writing of 'Athens', the opening of the door). For the example of my sister, he would consider that if I say 'My sister gave birth in 2005', I describe an event, but if I say 'I

became an uncle', I don't describe an event. (Perhaps he would tackle Peter's example in the same way, that is, by considering that his turning 21 in 2021 was not an event.) But then, the question arises for Lewis when some event corresponds to our words and when it seems to correspond but doesn't really.

(c) The problem of overdetermination. An event z was *overdetermined* iff there existed two events (or other entities), x and y, such that: x caused z, y also caused z, and x was not a cause of y, nor was y a cause of x. For example, A and B shoot C in the head from nearby and simultaneously, but from opposite directions; then x is A's shooting, y is B's shooting, and z is C's death.

If x hadn't occurred, z would have occurred because of y. And if y hadn't occurred, z would have occurred because of x. Indeed, it seems that, in such cases, there is no such series of events from x to z as is necessary, in Lewis's view, for there being causality. For when we reflect on the consecutive results of x on C's organism, however close we may get to z we do not find any event such that if it hadn't occurred, z wouldn't have occurred either. For the same reason, there seems to be no appropriate series of events from y to z. But of course x (just like y) is a cause of z.

The way Lewis tackled overdetermination changed over the years. But it is more interesting to see a way of tackling it that is not adopted by Lewis but ensues from some of Davidson's views (in his paper "The Individuation of Events") It may be right to say the following: If x hadn't occurred, z wouldn't have occurred. There would have occurred an event that was a death of C, but it wouldn't have been z. For z is a death from two bullets in the head. The event that would have been C's death if x hadn't occurred would have been a death from one bullet. So it wouldn't have been the same event. Similarly, if y hadn't occurred, again z wouldn't have occurred.

5. Causation and explanation:

We often try to explain an event, a phenomenon, a situation or something else. There are various kinds of explanation, but at any rate an explanation often has the form '**A** because **B**'. When we are explaining an event, we usually try to describe its causes. We sometimes have a correct explanation of the form '**A** because **B**' and both the sentence **A** and the sentence **B** correspond to an event, but the event that **B** corresponds to is not a cause of the event that **A** corresponds to. E.g. if I run a red light and say 'I violated the traffic code because I ran red', the running of the red light and the violation of the traffic code are the same event, which of course is not a cause of itself. (As Davidson has emphasized, one event can be described in various ways. E.g. the descriptions 'the assassination of Archduke Ferdinand' and 'the incident that the first world war began with' refer to the same event.)

We sometimes have a statement of the form '**A** because **B**', the event that the sentence **B** corresponds to is a cause of the event that **A** corresponds to, but the statement is not a correct explanation. For example, Oswald killed Kennedy. So let's suppose that he felt remorse. Let's also say that the assassination of Kennedy is the historical episode that has been shown on American TV more than anything else. In the statements 'Oswald was filled with remorse because he assassinated Kennedy', 'Oswald was filled with remorse because Kennedy was assassinated' and 'Oswald was filled with remorse because the historical episode occurred which has been shown on American TV more than anything else', the sentences that follow the 'because' correspond to the same event, and that event was a cause of the remorse. Still, only the first statement is a correct explanation.