Faculty of Philosophy, University of Cambridge
Michaelmas Term 2013

Part IA: Metaphysics
Causation

Brief description of lectures
These lectures will introduce the metaphysics of causation (or causality, or cause and effect). After a brief introduction to the topic, and some historical background, we will consider the concept of causal necessity and then look at David Hume's view of causation, followed by David Lewis's counterfactual theory, and the main objections to it.

Reading List (*Essential reading is marked* *; more difficult reading marked **)
Anthologies
Ernest Sosa, and Michael Tooley (eds.) *Causation* (Oxford: Oxford University Press 1993). An excellent anthology which contains many central readings. **Referred to below as S&T.**
Tim Crane & Katalin Farkas (eds.) *Metaphysics: a Guide and Anthology* (Oxford: Oxford University Press 2004). Section 6 contains an introduction to the topic and some essential readings on causation. **Referred to below as C&F.**

Lecture 1: The notion of cause
Introduction to section 6 of C&F
E.J. Lowe, *A Survey of Metaphysics* (Oxford: Oxford University Press 2001) chapters 8-10
*Jonathan Schaffer, 2003, 'The Metaphysics of Causation' *Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/entries/causation-metaphysics/
*RUSSELL, B., 'On the Notion of Cause', *Proceedings of the Aristotelian Society*, 13 (1913): 1-26. Reprinted in his *Mysticism and Logic* (London: Allen & Unwin, 1963)
**G.E.M. Anscombe, *Causality and Determination* (Cambridge: Cambridge University Press, 1971). Reprinted in P. van Inwagen and D. W. Zimmerman, eds., *Metaphysics: The Big Questions* (Oxford: Blackwell, 1998); also in S&T

2. Hume on causation and regularity theories
Tim Crane, 'Causation' in A.C. Grayling (ed.) *Philosophy: a Guide Through the Subject* (Oxford: Oxford University Press 1995).
David Hume, *A Treatise of Human Nature*, Book 1.iii., sects. 1, 2 & 14. Also available online at: www.gutenberg.org/ebooks/4705
*David Hume, *Enquiry Concerning Human Understanding* §IV, part I; and §VII, part II; reprinted in C&F. Also available online at: http://lib.myilibrary.com/?id=116084.
Jaegwon Kim, 'Causes and Events: Mackie on Causation', *The Journal of Philosophy*, 68 (1971): 426-41. Reprinted in S&T
*J.L. Mackie, 'Causes and Conditions' *American Philosophical Quarterly*, 2 (1965): 245-64. Reprinted in S&T

3. The counterfactual analysis of causation
* David Lewis, 'Causation', *Journal of Philosophy* 70 (1973): 556-67. Reprinted in Lewis's *Philosophical Papers Volume II* (New York: Oxford University Press 1986); and in S&T and C&F.
Paul Horwich, 'Lewis's Programme' in S&T
Jaegwon Kim, 'Causes and Counterfactuals' in S&T
*Peter Menzies, 'Counterfactual Theories of Causation' *Stanford Encyclopedia of Philosophy* http://plato.stanford.edu/entries/causation-counterfactual/

4. The problem of redundant causation
**David Lewis, 'postscripts' to 'Causation' in Lewis's *Philosophical Papers Volume II* (New York: Oxford University Press 1986).
*David Lewis, 'Void and Object' in Hall, N., & L. Paul, eds., *The Counterfactual Analysis of Causation* Cambridge, Mass: MIT Press 2004.
*Michael McDermott, 'Redundant Causation', *British Journal for the Philosophy of Science*, 40 (1995): 523-544.
**L.A. Paul and Ned Hall, *Causation: A User's Guide* (Oxford: OUP 2013), chapters 3 and 6
**Jonathan Schaffer 'Trumping Preemption', *Journal of Philosophy*, 97 (2000): 165-81

Faculty of Philosophy, University of Cambridge
Part IA: Metaphysics
Causation

Lecture 1: On the notion of cause

1.   The idea of cause
In the most general sense, to give the cause of something is to say why it is the way it is. When we ask, in the most general terms, why things are the way they are, we are asking about the causes of things. But the word 'cause' has been used for many different ideas in the history of philosophy'; these lectures will focus on one of them.

2.   Aristotle on causation
Aristotle distinguished between four ways in which something can be said to be a cause. Take a particular statue made by Michelangelo, e.g. a statue on the Medici tombs in Florence. Aristotle's view was that we can say that its cause is the marble from which it is made; or that its cause is the form or structure it has; or that its cause is what the sculptor who made it did when making it; or that its cause is the purpose for which it was made (e.g. to adorn a temple).

It might sound odd to talk of all these things as causes; in what sense, for example, is the 'form' of something a *cause*? However, if we understand the idea of 'giving a cause' broadly, in terms of giving an account of *why something is the way it is*, then it makes more sense. For each of the four answers above is, in a different sense, an answer to the question, 'why is the statue the way it is?'.

Since it is standard these days to describe explanations as 'answers to why-questions', it might be more natural to talk of four kinds of *explanation* here, rather than four *causes*. (Some commentators have talked in terms of Aristotle's 'four becauses' rather than 'four causes'.)

Aristotle's four causes (or 'four becauses') can be described as follows. About a particular statue, we might ask: 'why is this particular statue the way it is?'

Aristotle gave four answers:
(i)   The *material cause*: because it is made out of marble
(ii)   The *formal cause*: because it has the form of a statue
(iii)   The *efficient cause*: because Michelangelo made it
(iv)   The *final cause*: because it is needed to adorn the tomb

5.   **Efficient causes and effects: events?**
Since the 18th century, discussions of causation tend to be about what Aristotle called 'efficient causation'. The efficient cause of something explains why it happens, why it comes about.

Things that happen, or come about, are events. Events are concrete (i.e. spatio-temporal) particulars which have temporal parts. Objects, by contrast, are spatio-temporal particulars which do not have temporal parts. (NB This distinction is controversial! Some philosophers deny that objects and events should be distinguished in this way.)

So it is natural to say that all *effects* of efficient causes are events. But are all *causes* events? The way we normally talk about causation allow many kinds of things to be causes, other than events. It was Louis who broke the bottle; the bottle broke because it was fragile; it was the fact that he was so careless that caused this tragedy. But neither Louis, nor the fact that he was so careless, nor the bottle's fragility are events. One is an object, another a fact, and the other a property or feature.

None of these entities are events. But some philosophers (e.g. Davidson, Lewis) think that all causation must fundamentally relate events, and all other causation must be explained in terms of this. Why do they think this?

One reason for thinking that causes are events is that cause and effect must follow one another in time, and only events follow each other in time. Hume, for example, said that causes 'follow' effects. This must mean 'follow in time'. And this suggests, as Davidson ('Causal Relations') argued, that causes and effects are events, since events are the kind of entities which follow one another in time.

6.   Causal necessity
One event following another in time is not sufficient for the first to cause the second. Something could happen before something else without the first causing the second.

To think otherwise is to commit the fallacy known as *post hoc ergo propter hoc* ('after this therefore because of this').

Suppose there is a fire in a house and that a man was smoking a cigarette inside the house before the fire; the mere fact that this occurred before the fire is not enough to establish it as the cause. The cause must be the thing which *produced* or *brought about* the effect. But 'producing' and 'bringing about' seem to be synonyms for 'cause' here, so they can give an analysis. So what more should we say?

This is where philosophers often bring in the idea of *necessitation*: what *must* happen, given other things. The basic idea is that if the cause happened, then (in some sense) the effect must happen.

But what kind of necessity is this? Suppose it was the case that the sinister intruder tossed his cigarette onto a pile of newspapers. Then we can conclude that, given the laws of nature (= the general ways our world is), the newspapers *must* burn. But this 'causal necessity' does not seem to be the necessity of logic: there is no logical contradiction in asserting both 'The intruder threw his cigarette onto the newspapers' and 'The newspapers did not burn'. So if there is a necessity here, it seems to lie in the way the natural world is, and not in logic. (NB Logical necessities are those necessities which are due to logic. There may be necessities – necessary truths, things which are true in all possible worlds – which are not a matter of logic. 'Nothing can be red and green all over' is an example of such a necessity.)

7.   Hume on causation: an introduction

Some philosophers have thought that 'causal necessity' or 'natural necessity' is a very strange thing. To explain causation in terms of natural necessity might look like explaining one obscure idea in terms of another. This is what David Hume thought.

Hume thought that our idea of causation involved the idea of necessary connection, but that this was ultimately a confusion in our thinking about the world. For Hume, every idea must derive from a corresponding impression (experience). But he argues that we have no impression of necessary connection ('All events seem entirely loose and separate' he writes, 'one event follows another; but we can never observe any tie between them. They seem *conjoined*, but never *connected*'), and that therefore we have no adequate idea of it either.

Our idea of causation rather involves two other ideas: (i) temporal succession; and (ii) constant conjunction. Hence his definition: 'a cause is an object, followed by another, where all objects similar to the first are followed by objects similar to the second'.

Lecture 1: Hume and regularity theories of causation

1. Hume on causation

Hume is the originator of the regularity analysis of causation. Causation for Hume is defined in terms of two ideas: (a) temporal succession and (b) constant conjunction.

It is constant conjunction that is the basis in experience (and in reality) for our idea of necessary connection. Notice that constant conjunction only applies to kinds of events: we can't say that individual events are 'constantly conjoined' (we can't say that the assassination of the Archduke Franz Ferdinand is *constantly conjoined* with the first world war!). We have to say that events of *kind* A are conjoined with events of *kind* B. (Question: how would Hume explain an apparent causal truth like, 'the assassination of the Archduke was a cause of the first world war'?)

A questions about Hume: is he saying that there is no such thing as causation (an 'eliminative' view) or that constant conjunction and temporal succession is what causation really is (a 'reductive' view)?

2. Regularities and laws

The existence of a regularity between an event e of kind A and an event f of kind B is not sufficient for e to be a cause of f. My trips to the countryside (kind A) are invariably followed by rain (kind B). But this does not mean that a particular trip e is the cause of a particular rain shower f. We might try and escape this by distinguishing between those regularities which are *laws of nature* and those which are not (so-called 'accidental' regularities). So the kinds in question must be kinds which figure in laws of nature: kinds like 'having a mass of *n* grams'. Hence regularity views are sometimes called 'nomological' views of causation, *nomos* being the Greek word for law. (Question: how would ordinary apparent causal truths like the one above be understood on a nomological view?)

The regularity view explains what it is for a particular event to cause another in terms of truths about *other events*. The causal relation is a purely local relation relating *this* cause c and *this* effect e: rather the fact that this relation holds implies that relations hold between all sorts of other events of the same kinds. Is there anything more we can say about how *particular causes* are related to *particular effects* apart from the fact that they are instances of laws?

3. Causes and conditions

JL Mackie ('Causes and Conditions') tries to characterise in more detail how a cause is related to its effects in terms of a combination of necessary and sufficient conditions. Mackie argues that a cause is not sufficient for its effects, and nor is it necessary. Take Mackie's example: a short circuit in a house causes a fire. The short circuit is not sufficient (=enough) because without the presence of inflammable material and oxygen (etc.) the short circuit would not have brought about the fire. But nor is it necessary, Mackie says, since something other than the short circuit could have caused the fire. Rather, he claims, that the short circuit is an **I**nsufficient but **N**ecessary part of a condition which it itself **U**nnecessary (for the effect) but **S**ufficient. The cause is what he calls an **INUS** condition. The sufficient condition is a very large collection of conditions which absolutely suffices for the effect. (This is in the spirit of Mill, who said that the 'whole cause' is the 'entire state of the universe' before the effect.)

We can simplify Mackie's analysis. Is the cause sufficient? Mackie says: yes, given the other conditions or circumstances. Is the cause necessary? Yes, given the other conditions or circumstances. Illustration: given the oxygen, the inflammable material (etc.) the short circuit is sufficient for the fire; given the oxygen, the inflammable material (etc.), the short circuit is necessary for the fire. So maybe we can just say: the cause is necessary and sufficient *in the circumstances* for its effects.

However, there is a problem with saying that a cause is sufficient for its effects; it commits us to all causation being deterministic. (For a definition of determinism, see Lewis 'Causation'.) But is all causation deterministic? Modern physics suggests not. We should not, perhaps, commit ourselves to the idea that causes are sufficient, even sufficient in the circumstances.

What about the cause as a necessary condition for its effects? The question is, what is it for a cause to be 'necessary in the circumstances' for its effects, how should this be understood? We will explain this when we move on to the counterfactual analysis of causation. Before moving on we have to clarify one issue which arises from Mackie's theory: in what sense we can talk about 'the cause' of an effect.

4.   Every effect has many causes

One obvious consequence of Mackie's analysis is that every event has many causes. For just as the short circuit is sufficient and/or necessary for its effects given the other circumstances (including the presence of oxygen), so the presence of oxygen is sufficient and/or necessary for its effects, given the other circumstances (including the short circuit). And we can say the same about every other INUS condition of any effect.

Some philosophers resist the idea that things other than events (like the short circuit) are causes, and say that they are only the 'circumstances' or 'background conditions'. They then attempt to produce a theory of causation which will distinguish causes from circumstances. But it is very hard to apply these definitions to actual cases without stretching the facts to fit the theory, or making *ad hoc* stipulations. When we talk about things like '*the* cause of Caesar's death', this is best regarded as loose talk. Caesar died because his heart stopped; but he also died because he was stabbed; if he had been as strong as an ox maybe he would have been able to escape his assassins; and maybe he also died because he was ambitious. What general reason is there to think that every event has only one immediate cause? We are perfectly happy with the idea that an effect may have had many causes stretching back in time (many 'mediate' causes). So surely it makes just as much sense to say that there many causes of an effect immediately before it happened. The truth is surely that every effect has many causes; so it cannot be an objection to the regularity theory that it cannot distinguish causes from 'circumstances' or 'background conditions'.

5.   Sample questions

(i) Are causes necessary, and/or sufficient for its effects? (ii) Is there a real distinction between cause and background circumstances? If not, what is the explanation for the illusion that there is? (iii) In what sense, if any, do events have many causes?

Lecture 3: The Counterfactual Analysis of Causation

1. 'The thing that made the difference'

Recall that an event can be sufficient in the circumstances for the occurrence of an effect, but not be a cause of that effect. A plausible idea instead is that the cause is the think that makes the difference to whether the effect happens or not. This is the idea Lewis uses to define causation, along the lines of Hume's 'where the first object had not been, the second never had existed'.

2. **Counterfactual dependence**

Lewis defines causation in terms of the truth of certain counterfactual conditionals. A conditional is any 'if… then…' statement or sentence. The sentence coming after the 'if' is called the antecedent, and the sentence coming after the 'then' is called the consequent. A *counterfactual* conditional is a conditional of the form 'if it were the case that A, then it would be the case that C'. (Lewis symbolises this as 'A $\square\rightarrow$ C'.)

Normally, we think of counterfactual conditionals (from now on simply call them 'counterfactuals') as having antecedents which are 'contrary to fact' (that is not true). But this need not be. Just think instead of counterfactuals as statements of the form 'if it were the case that A, then it would be the case that C'.

The key notion which Lewis employs in his account of the truth-conditions of counterfactuals is the notion of *comparative overall similarity of possible worlds*. We shall look more at this notion in the future, but for the moment take it in an intuitive way: one world is more or less similar to another to the extent that it resembles that world in matters of particular fact and law. In what follows, an 'A-world' is a world in which A is true; likewise with C.

Lewis defines the truth-conditions of counterfactuals as follows:
'A $\square\rightarrow$ C' is true iff either (1) there are no possible A-worlds (i.e. it is impossible that A); or (2) some A-world which is also a C-world is closer to the actual world (actuality) than any A-world which is not also a C-world.

(1) is the situation in which the counterfactual is said to be 'vacuously true'. The important case is (2). The idea is that when the counterfactual is true it takes less of a departure from actuality to make A and C true together than it does to make A true without C. Try applying this idea to an example (e.g. Lewis's 'if it were the case that kangaroos had no tails, then it would be the case that they would topple over').

When 'A $\square\rightarrow$ C' is true we can say that C counterfactually depends on A.

3. Causal dependence

Using the notion of counterfactual dependence, which is a relation between propositions, Lewis defines *causal* dependence, a relation between events. Where 'c' and 'e' are terms denoting events (e.g. 'the assassination', 'the first world war') and 'O' is a predicate of events, meaning 'occurs', and '~' is negation, we can define causal dependence as follows:

e causally depends on c iff:
(1) Oc $\square\rightarrow$ Oe; and
(2) ~Oc $\square\rightarrow$ ~Oe

If c and e are actual events, then (1) is automatically true because of the stipulation that the actual world is always the closest world to itself. Since c and e actually exist, then there is a c-and-e world which is closer to actuality than any c-and-not-e world, simply because the actual world *is* a c-and-e world. And in any case of causation, the cause and the effect must actually exist. The clause (2) is the one which is normally taken to be the heart of the counterfactual analysis: it says that if e had not occurred, e would not have occurred. This is what it is for e to causally depend on c.

But there are many cases of counterfactual dependence which are not cases of causation (see Kim's paper in the Sosa & Tooley volume for some examples). These can even be cases of counterfactual dependence between events. If I park my car on a double yellow line, then I break the law. But parking the car on double yellow lines does not *cause* me to break the law, rather it constitutes breaking the law in this instance. What we have to have to add to exclude cases like this is to say that the events related as cause and effect must be *distinct* from one another.

4. **Sample questions**

(i) Explain and illustrate the difference between the concept of causal dependence and the concept of causation. (ii) Similarity of worlds (iii) Explain why causation is transitive and causal dependence is not.

Lecture 4: The problem of redundant causation

5. Recap: counterfactual dependence and causal dependence

Lewis defines causation in terms of two ideas: (i) counterfactual dependence (a relation between propositions); and (ii) causal dependence (a relationship between events). Last week we looked at the notion of counterfactual dependence. Recall that:

'A $\square\rightarrow$ C' is true iff either
(1) there are no possible A-worlds (i.e. it is impossible that A); or
(2) some A-world which is also a C-world is closer to the actual world (actuality) than any A-world which is not also a C-world.

and when 'A $\square\rightarrow$C' is true, we say that C counterfactually depends on A.

Why not define causation simply in terms of counterfactual dependence? Because there are many cases of counterfactual dependence which are not cases of causation (see Kim's paper in the Sosa & Tooley volume for some examples).

There can even be cases of counterfactual dependence between events which are not cases of causation. If I park my car on a double yellow line, then I break the law. But parking the car on double yellow lines does not *cause* me to break the law, rather it constitutes breaking the law in this situation. What we have to have to add to exclude cases like this is to say that the events related as cause and effect must be *distinct* from one another.

Using the notion of counterfactual dependence, which is a relation between propositions, Lewis defines *causal* dependence, a relation between events. Where 'c' and 'e' are terms denoting events (e.g. 'the assassination', 'the first world war') and 'O' is a predicate of events, meaning 'occurs', and '~' is negation, we can define causal dependence as follows:

e causally depends on c iff:
(1) Oc $\square\rightarrow$ Oe; and
(2) ~Oc $\square\rightarrow$ ~Oe

If c and e are actual events, then (1) is automatically true because of the stipulation that the actual world is always the closest world to itself. Since c and e actually exist, then there is a c-and-e world which is closer to actuality than any c-and-not-e world, simply because the actual world *is* a c-and-e world. And in any case of causation, the cause and the effect must actually exist.

The clause (2) is the one which is normally taken to be the heart of the counterfactual analysis: it says that if c had not occurred, e would not have occurred. This is what it is for e to causally depend on c.

6. Causation defined

Lewis says that causal dependence between actual events implies causation, but not vice versa. This is because causation is a transitive relation and causal dependence is not.

A relation R is transitive when 'aRb' and 'bRc' imply 'aRc'. A relation is non-transitive when this is not the case; a relation is *intransitive* when 'aRb' and 'bRc' imply 'not-aRb'.) Rather, causation is defined in terms of a *chain* of counterfactual dependence. A *causal chain* is defined as a sequence of actual events, c, d, e… etc., where d depends on c, e depends on d etc. Then c is a *cause* of e when there is a causal chain from c to e.

Lewis's definitions allow that there could be a sequence where *a* causally depends on *b*, *b* causally depends on *c*, but *a* does not causally depend on *c*. Nonetheless, Lewis says, it will still be true that in this case, *c* is a cause of *a*.

Thus suppose I shoot the president, and this brings about a revolution, which in turn brings about the president's rival ascending to power. Let's suppose that each later stage in this process causally depends on the earlier stage. Lewis would say that even though it is true that my act caused the president's rival to ascend to power, it need not thereby be true that the president's rival's ascent is causally dependent on my shooting, since it need not be true that in the closest world in which I did not shoot, he did not ascend to power (maybe the whole situation is so politically unstable that someone else would have shot if I hadn't). So we have causation between my action and the eventual outcome without causal dependence.

Lewis therefore defines causation as the *ancestral* of the relation of causal dependence. The ancestral of a relation R is that relation which stands to R as the relation of *being an ancestor* stands to the relation of *being a parent*. The relation 'ancestor' can be roughly defined as follows: x is an ancestor of y iff x is a parent of y, or x is a parent of a parent of y, or x is a parent of a parent of a parent of y... and so on. So while 'x is a parent of y' is not transitive, 'x is an ancestor of y' is. The same structure holds for the relations 'x causally depends on y' and 'x is a cause of y'.

7. The problem of redundant causation

Apparent cases of redundant causation occur when it is obvious that c caused e but it is not true that if c had not occurred, e would not have occurred. Two kinds of case are typically considered:

(i) Overdetermination: e has two independent causes, c1 and c2. Since c1 and c2 are independent, either would occur without the other. In the closest world where c1 does not occur, e still occurs, because c2 brings it about. And vice-versa with c2. So e causally depends on neither c1 nor c2.
Example: two independent assassins shoot and kill the tyrant. Arguably in such a case, if the first assassin had not fired, the second would still have killed the tyrant. And the same applies to the second.

(ii) Pre-emption: c causes e, but there is a 'back-up' cause waiting to cause e if c fails. So e does not causally depend on c.
Example: one assassin shoots and kills the tyrant. But another is waiting to shoot just in case the first one misses. The first one doesn't miss, so his shot is the cause of the death. But it isn't true that if the first had missed, the tyrant would not have died.

Lewis deals with pre-emption by appealing to his definition of causation in terms of a chain of causally dependent events. The first assassin's shot causes the tyrant's death because there is a chain of events between this shooting and the death. Each event in this chain is causally dependent on the one before, but this does not mean that the death is causally dependent on the shooting. So the fact that there is a back-up does not stop the shot causing the death.

But what if there is no chain? What if the pre-empting cause occurs immediately before the pre-empted one? This has become known as 'late pre-emption', and Lewis cannot rule it out as impossible. See the readings by Schaffer, Menzies and Hall & Paul for discussion.

8. Sample questions

(i) Explain and illustrate the difference between ovdetermination and pre-emption. Why does this difference matter for theories of causation? (ii) Why can't Lewis reject these cases on the grounds that if the effect had had a different cause, it would have been a different effect? (See Lewis, 'Postscripts' to 'Causation', on the reading list.)

Tim Crane
November/December 2013
www.timcrane.com