

## The Concept of Causation— Conditional Analyses

ONE of the many problems with which Hume's discussion leaves us is this: what is our concept of causal as opposed to non-causal sequences; what do we (rightly or wrongly) take to be their distinguishing feature? The obvious, naïve, way of tackling this question would be to compare two otherwise similar sequences, only one of which we recognize as causal, and see what distinctive comment we should make about the causal one. Let us, then, compare the following sequences A and B.

A: A chestnut is stationary on a flat stone. I swing a hammer down so that it strikes the chestnut directly from above. The chestnut becomes distinctly flatter than before.

B: A chestnut is stationary on a hot sheet of iron. I swing a hammer down so that it strikes the chestnut directly from above. At the very instant that the hammer touches it, the chestnut explodes with a loud pop and its fragments are scattered around.

I assume that we know or perceive enough about sequences A and B to be able to say that the chestnut's becoming flatter in A was caused by its being hit with the hammer, but that its exploding in B was not caused by its being hit with the hammer. (Among other things, the explosion was not due to the hammer's pushing the chestnut into closer contact with the hot metal, or anything like that.) I am not concerned, at the moment, with how we know this, but merely with what we mean, what we are *saying* when we say that the blow-flattening sequence was causal but that the blow-exploding sequence was not.

I suggest that the obvious answer, the one that pretty well

any intelligent but unprejudiced person would give to this question, is that in A the chestnut would not have become flatter if the hammer had not struck it, whereas in B it would have exploded even if the hammer had not struck it. It is tempting to say, then, that the contrast between these two contrary-to-fact conditionals is at least an initial indication of what we mean by calling the one sequence causal and the other not, that is, that to ascribe what we called necessity<sub>1</sub> is to assert such a contrary-to-fact conditional as the former. As Mill says, 'if a person eats of a particular dish, and dies in consequence, that is, would not have died if he had not eaten of it, people would be apt to say that eating of that dish was the cause of his death'.<sup>1</sup> They would indeed.

Curiously enough, Hume says this too. In the *Enquiry*, after giving the first, regularity, definition ('we may define a cause to be an *object*, followed by another, and where all objects similar to the first are followed by objects similar to the second'), he adds 'or in other words where, if the first object had not been, the second never had existed'.<sup>2</sup> Now of course this variant is *not* the regularity definition in other words. (It does not even follow from the regularity definition as given here, though on a certain interpretation of the conditional it might be held to follow from, and even to be equivalent to, a different regularity statement, namely 'where all the objects similar to the second are preceded by objects similar to the first'). But I shall not spend time on the questions why Hume asserted this implausible equivalence, or why he did not realize that this was a much better answer than the one he had given in all his previous discussion to the question 'What is our *idea* of necessary connection?', but consider directly the merits and the defects of this suggestion.

To simplify matters, I shall for the present be considering only singular causal statements, especially statements about individual event sequences, such as 'The hammer's striking the chestnut caused the flattening of the chestnut'. I shall speak of causes and effects (or results) as events, but 'event' is just a general term that stands in for such items as 'the hammer's

<sup>1</sup> *System of Logic*, Book III, Ch. 5, Sect. 3. Conditional analyses of causal statements are defended and discussed by Ardon Lyon in 'Causality', *British Journal for the Philosophy of Science*, xviii (1967), 1-20.

<sup>2</sup> *Enquiry*, Sect. VII, Pt. II, Selby-Bigge, p. 76.

striking the chestnut'. (Anxieties about the exact ontology of causation may be postponed for later consideration in Chapter 10.) Such phrases are nominalizations of singular event-sentences or event-clauses; e.g. 'the flattening of the chestnut' is a nominalization of 'the chestnut became flatter'. The present suggestion, then, is that a statement of the form '*X* caused *T*' means '*X* occurred and *T* occurred and *T* would not have occurred if *X* had not', it being understood that when we instantiate '*X*' and '*T*' with particular event descriptions we can express the suggested meaning most neatly by going back from the nominalizations to the corresponding clauses or their negations: e.g. 'the striking of the match caused the appearance of the flame' would, on this suggestion, mean 'the match was struck and the flame appeared and the flame would not have appeared if the match had not been struck'.

But is the suggestion correct?

First, are there cases where we would say that *X* caused *T* but would not say what is proposed above as the meaning of this claim? Well, clearly we cannot say that *X* caused *T* unless both *X* and *T* occurred, provided that we take 'occurred' in a fairly broad sense, to include the presence or persistence of standing conditions. But what about the counterfactual conditional? Might we not say that the striking of the match caused the appearance of the flame and yet admit that even if the match had not been struck the flame would have appeared if, say, this match had been touched by a red-hot poker? Certainly we might, so the suggestion needs some modification at least: let us insert 'in the circumstances', reading '. . . and in the circumstances the flame would not have appeared if the match had not been struck'. 'The circumstances' can be taken to include the fact that this match was not in fact touched by a red-hot poker at the critical time. And in general we can modify our suggestion to read '. . . in the circumstances *T* would not have occurred if *X* had not', the qualifying phrase being interpreted in the sort of way indicated by our example. An objector, however, might say, 'But what if, for example, the match really was touched by a red-hot poker at the same instant that it was struck? Then even *in the circumstances* the flame would have appeared even if the match had not been struck.' True, but it is not clear whether in this case we would say that the striking

of the match caused the appearance of the flame. What we would or should say in such cases of causal over-determination, of fail-safe causes, will be considered later: for the moment we need not take them as falsifying our modified suggestion.

Secondly, and more awkwardly, are there cases where we would *not* say that *X* caused *Y* but would say that *X* and *Y* both occurred and that in the circumstances *Y* would not have occurred if *X* had not? Would we not have to say the latter, trivially, if *X* and *Y* were identical? Provided that *X* occurred, we must admit that *X* occurred and *X* occurred and in the circumstances *X* would not have occurred if *X* had not. But events, unlike Spinoza's God, are not commonly said to cause themselves. Equally, the penny would not have fallen heads-up if it had not fallen tails-down; but we would not say that its falling tails-down caused its falling heads-up. Again, the driver would not in the circumstances have broken the law if he had not exceeded the speed limit; but we would not say that his exceeding the speed limit caused his breaking of the law. To exclude all such counter-examples, we must say that, in addition to the meaning suggested above, '*X* caused *Y*' presupposes that *X* and *Y* are distinct events, and, as the last two examples show, this must be taken in a stronger sense than merely that '*X*' and '*Y*' are logically or conceptually independent descriptions. Indeed, it is not even necessary, any more than it is sufficient for this purpose, that these should be logically independent descriptions. It is legitimate, though trivial, to say that *X* caused every effect of *X*; again I can say that my meeting Tom on Monday caused my remembering, on Tuesday, my meeting Tom on Monday. And so on. Logically independent descriptions, then, are not called for; what is required is that the cause and the effect should be, as Hume says, distinct existences. It may be objected that this requirement is vague or obscure, but it is not, I think, necessary for me to aim at any great precision here. I am discussing only what causal statements mean, and for this purpose it is sufficient to say that someone will not be willing to say that *X* caused *Y* unless *he* regards *X* and *Y* as distinct existences. But are there really no limits to the possible descriptions of *X* and *Y*? What about *irrelevant* descriptions, ones that have, as we would say, nothing to do with the causal relationship? Some seem acceptable: 'the

incident in the tennis court advanced the French Revolution.' But consider 'Jane's eating fishpaste in the pantry caused her feeling ill' and the corresponding suggested expansion 'Jane ate fishpaste in the pantry and felt ill and in the circumstances she wouldn't have felt ill if she hadn't eaten fishpaste in the pantry'. We are inclined to comment that she felt ill either because she ate too much fishpaste or because the fishpaste was bad, and in either case it wouldn't have mattered where she ate it. This raises the question (which will be discussed in Chapter 10) about the extensionality of causal statements. For the present it is enough to say that the issue is in doubt: we might want to reject or deny both the proposed singular causal statement and its expansion, or we might want to say that so long as it was eating the fishpaste that made Jane ill, and it was in the pantry that she ate it, they are both true but merely misleading. But it seems that whatever we decide to say, we can say about both the singular causal statement and its proposed expansion, so that examples of this sort do not tell against the suggestion that the latter shows what the former means. Cases where  $X$  and  $Y$  are collateral effects of a common cause create a more serious difficulty. Labour's defeat at the election pleases James but saddens John, who, as it happens, are quite unknown to each other. Then James's being pleased does not cause John's being sad, and yet we might well say that in the circumstances John would not have been sad if James had not been pleased. Yet there is a way of handling the words 'if' and 'in the circumstances' that will defeat this supposed counter-example. Construct a 'possible world' in the following way: take something that is just like the actual world up to the point where, in the actual world, James is pleased; keep out, from your possible world, James's being pleased, but otherwise let your possible world run on in the same manner as the actual world; only if John does not become sad in your possible world can you say 'in the circumstances John would not have been sad if James had not been pleased'. And if you *can* say this, then you can also say that James's being pleased caused John's being sad (in the actual world). But this special way of handling 'if in the circumstances' runs a grave risk of circularity if we want to use this conditional as an analysis of causal statements, since in explaining this handling we have had to use the notion of

letting the possible world run on. Yet without this special interpretation of the counterfactual conditional, it seems that while it may be required for the truth of the corresponding causal statement, it is not sufficient for this.

Another difficulty is concerned with the distinction which we are inclined to draw between conditions and causes. There may be a set of factors which were, in the circumstances, jointly sufficient and severally necessary for a certain result, and which all occurred, as, consequently, did the result. Then we can say of each of these factors that if in the circumstances it had not occurred the result would not; but we may not be so willing to say of each of them that it caused the result. Perhaps we *ought* to say this, since they are all logically related to the result in the same way: the situation as described is symmetrical with respect to all the factors. But at present we are discussing not what we ought to say but what our causal statements mean. And there is no doubt that we tend to be a bit selective, to be more ready to call some kinds of factors causes than others. There are no firm rules governing this selection, but there are some fairly systematic tendencies.

Thus we are more ready to say that an event caused a certain effect than that a standing condition did: it was the spark rather than the presence of inflammable material that caused the fire. Even among events ones which are seen as intrusive are picked as causes in preference to ones which occur within some going concern: it was the severing of the artery rather than the pumping of the heart that caused the loss of blood. This preference shades into one governed by moral and legal assumptions: what is normal, right, and proper is not so readily called a cause as is something abnormal or wrong. This third tendency may conflict with and override the first. Since it is normal for people to be striking matches and lighting cigarettes in a residential flat, but a gas leak is abnormal and should not occur, we may well say that the explosion which wrecked this block of flats was caused by the presence of a quantity of gas rather than that it was caused by Jones lighting his cigarette.

These matters may be clarified to some extent if we realize that causal statements are commonly made in some context, against a background which includes the assumption of some

*causal field*.<sup>3</sup> A causal statement will be the answer to a causal question, and the question 'What caused this explosion?' can be expanded into 'What made the difference between those times, or those cases, within a certain range, in which no such explosion occurred, and this case in which an explosion did occur?' Both cause and effect are seen as differences within a field; anything that is part of the assumed (but commonly unstated) description of the field itself will, then, be automatically ruled out as a candidate for the role of cause. Consequently if we take the field as being *this block of flats as normally used and lived in*, we must take Jones's striking a match to light his cigarette as part of the field, and therefore not as the cause of, or even a cause of, or as causing, the explosion. What caused the explosion must be a difference in relation to the field, and the gas leak, or the presence of the gas that had leaked out, is the obvious candidate.

What is said to be caused, then, is not just an event, but an event-in-a-certain-field, and some 'conditions' can be set aside as not causing this-event-in-this-field simply because they are part of the chosen field, though if a different field were chosen, in other words if a different causal question were being asked, one of those conditions might well be said to cause this-event-in-that-other-field. Any part of the chosen field is decisively ruled out as a cause; a more elusive point is that among factors not so ruled out which are still severally necessary for the effect, we still show some degree of preference on the grounds indicated above. But I think that this can be taken as reflecting not the meaning of causal statements but rather their conversational point, the sorts of use to which they are likely to be put. We often want to know what caused some event with a view to saying how it could, and perhaps should, have been prevented: pointing to the spark rather than to the presence of inflammable material indicates what would have been the last chance of preventing the fire. But it might be conceded that the statement that the presence of this material caused the fire would be

<sup>3</sup> This notion of a causal field was introduced by John Anderson in 'The Problem of Causality', *Australasian Journal of Psychology and Philosophy*, xvi (1938), reprinted in his *Studies in Empirical Philosophy*, and used to resolve difficulties in Mill's account of causation. I also used it in 'Responsibility and Language', *Australasian Journal of Philosophy*, xxxiii (1955), 143-59, to deal with problems of moral and legal responsibility.

as true as the statement that the spark caused it, and merely in some ways less interesting. We can and do indeed distinguish between triggering causes and predisposing causes, which shows that standing conditions are not prevented from being causes by the mere meaning of the noun, or of the verb, 'cause'. Similarly we may agree that the collision was caused just as much by Smith's driving straight ahead as by Brown's deviating to his right without warning, but say that it is more important, for moral and legal purposes, to draw attention to the second of the two causal relationships.

The supposed distinction between conditions and causes can be adequately accounted for in these two ways: an alleged condition which is not called a cause, although if in the circumstances it had not occurred the result would not, either is part of the field presupposed in the view taken by the speaker of the result (and so is not a cause in relation to this field) or is a cause, but mention of this fact happens to be irrelevant, or less relevant than mention of some other cause of the same result, to some current purpose.

There is admittedly some logical redundancy in the two treatments offered here of conditions which we are reluctant to call causes. We should get a neater account if, say, we assigned all such conditions to the assumed causal field, and said that anyone who withheld the title of cause from what he admitted to be one of a set of severally necessary and jointly sufficient conditions for some result was implicitly taking that condition to be part of the field. Again, we should get a neater account if we discarded the notion of a field, and interpreted the withholding of the title of cause from any such necessary factor as reflecting some conversational or other purpose of the speaker. But at present my object is not so much to give a logically neat account as to analyse our ordinary thinking about causal sequences, and I believe that this does contain the two separate elements I have tried to describe. On the one hand there is the implicit question, 'What caused such-and-such?', where 'such-and-such' is already thought of as a change in a certain material or background state of affairs; on the other hand, even among differentiae within this background, we tend to downgrade some and at least initially deny that they caused the result, though we might reluctantly concede, if pressed, that they helped to cause it.



Another sort of apparent counter-example is easily disposed of in the same way. It can be truly said of anyone now dead that if he had not been born he would not have died, and although it would here be pointless to insert 'in the circumstances', such an insertion would not make the statement false. Yet we would not say that being born caused his death—that is, not ordinarily: with sufficient ingenuity one can construct a case for saying this. But the reason is merely that when we look for a cause of someone's death, the event, this person's death, is a change in a field which centrally includes this person's being alive for a while, and hence (in the ordinary course of nature) his having been born. As before, his being born is part of the field, and therefore cannot be the cause of his death.

The suggested meaning for '*X* caused *Y*' which we have been considering is '*X* occurred and *Y* occurred and in the circumstances, *Y* would not have occurred if *X* had not'. A slightly different formula would run '. . . and there were circumstances such that in them *Y* would not have occurred if *X* had not'. To see the difference, suppose that the effect is that a certain part of a certain house catches fire, that the presence of oxygen, the absence of any extinguishing device, etc., are taken as parts of the field, and that in relation to this field the following three conditions, which were all present on the occasion in question, were severally necessary and jointly sufficient for the effect: an electric current in a certain wire (*A*), decayed insulation at a point on that wire (*B*), and inflammable material near that point (*C*). Then there were circumstances, namely *B* and *C*, such that in them the fire would not have occurred if the current, *A*, had not; equally there were circumstances, namely *A* and *C*, such that in them the fire would not have occurred if the faulty insulation, *B*, had not, and again circumstances, namely *A* and *B*, such that in them the fire would not have occurred if the inflammable material, *C*, had not. So by our modified formula, *A*, *B*, and *C* would each equally count as causing the fire. But someone who says that, for example, the faulty insulation, *B*, caused the fire, may be thinking of *A* and *C* as 'the circumstances', may be presupposing them rather than saying *that there are* circumstances in relation to which *B* was necessary. Before he will agree that *A* caused the fire, he will have to make

a switch of presuppositions, and take *B* and *C*, rather than *A* and *C*, as 'the circumstances'. So whereas with our modified formula there is no conflict at all between the three statements that *A* caused the fire, that *B* did so, and that *C* did so—since the three existential statements are entirely compatible—there is, with our first formula, using the phrase 'in the circumstances', a contrast of approach and presupposition between those who assert these three statements, although no outright disagreement. And this surely is the case. If three speakers put forward these three alternative causal statements, they would be conscious of such a contrast of approach. I think, therefore, that our first formula is the better for catching the force of a causal statement in use. The modified formula might be said to yield more clearly marked truth conditions for causal statements; but I shall argue that what it yields are not, strictly speaking, truth (or falsity) conditions. Still, it may be conceded that the modified formula represents a perhaps desirable tidying up of the meaning of causal statements, whereas the original formula comes closer to displaying the meaning they actually have.

There is, however, something surprising in our suggestion that '*X* caused *Y*' means, even mainly, that *X* was *necessary* in the circumstances for *Y*. Would it not be at least as plausible to suggest that it means that *X* was *sufficient* in the circumstances for *Y*? Or perhaps that it was both necessary and sufficient in the circumstances? After all, it is tempting to paraphrase '*X* caused *Y*' with '*X* necessitated *Y*' or '*X* ensured *Y*', and this would chime in with some of the thought behind the phrase 'necessary connection'. But if '*X* necessitated *Y*' is taken literally, it says that *Y* was made necessary by *X* or became necessary in view of *X*, and this would mean that *X* was sufficient rather than necessary for *Y*. Of course, an *X* which in ordinary discourse is said to cause a *Y* is practically never *in itself* sufficient for that *Y*, or even believed to be so; as before, we must add the qualification 'in the circumstances'. And then the suggested description seems to fit what we recognize as causes. Taking our last example again, and letting the circumstances include the electric current and the inflammable material, we can say that the faulty insulation (which on these assumptions caused the fire) was sufficient in the circumstances for it. And similarly in sequence A in our earlier example the hammer-

blow was sufficient in the circumstances for the chestnut's becoming flatter. But is sufficiency sufficient, or is necessity necessary as well? Can we explain the fact that in sequence B the hammer-blow did not cause the explosion on the ground that it was not sufficient for it in the circumstances? In fact the circumstances in themselves, without the hammer-blow, were sufficient: does it not follow that the hammer-blow was trivially and automatically sufficient in the circumstances, since anything at all, or nothing, would have been so, provided that the relevant circumstances were not tampered with? If we argued that what is only thus trivially sufficient in the circumstances cannot be properly so described, and that that is why it cannot be said to cause the result, are we not covertly reintroducing the requirement that a cause should be necessary in the circumstances rather than sufficient? What there is any point in mentioning as sufficient in the circumstances will be necessary in the circumstances as well.

This point seems to be confirmed if we try to expand the phrase 'sufficient in the circumstances' into a conditional statement. 'X will be sufficient in the circumstances for Y' may be taken as saying 'Given the circumstances, if X occurs then Y will', and even as a non-material conditional this will count as true provided that the circumstances do not change, X occurs, and Y occurs also.<sup>4</sup> And then 'X was sufficient in the circumstances for Y' will be equivalent to 'Given the circumstances, if X occurred, then Y did', and provided that the circumstances referred to are the actual ones this will automatically be true of any sequence in which X and Y actually occurred. Sufficiency in the circumstances is, then, of no use for our present purpose of finding the distinguishing feature of causal sequences; every cause is sufficient in the circumstances for its effect, but so are many non-causes for events which are not their effects.

But this is a weak sense of 'sufficient in the circumstances'. Can we find also a strong, counterfactual sense which will not apply to the antecedent in every actual sequence? This would be that, given the circumstances, if Y had not been going to occur, X would not have occurred. This is a possible sense, and what we recognize as causes are in general sufficient in the

<sup>4</sup> I have discussed non-material conditionals, both open and counterfactual, and their truth conditions, in Chapter 3 of *Truth, Probability, and Paradox*.

circumstances in this strong sense as well as in the weak one. In the appropriate possible world in which the circumstances are the same as in sequence A, but the chestnut does not become flatter, the hammer-blow has not occurred. (In constructing this possible world, we are, of course, taking over laws of working from the actual world.) And we cannot say the corresponding thing about the non-causal sequence B. The statement 'If in the circumstances of sequence B the explosion had not been going to occur, the hammer-blow would not have occurred' is not true or even acceptable. The supposition implicit in its antecedent cannot be coherently considered in the light of the actual world's laws of working, for, given these laws, in the circumstances of sequence B the explosion was going to occur. And if we take the antecedent as inviting us to consider some change in those laws of working, there is still no reason why we should not combine, with whatever coherent interpretation we place on this supposition, the view that the hammer-blow still occurred. It looks, then, as if the strong counterfactual sense of 'sufficient in the circumstances' will distinguish causal from non-causal sequences, though the weak sense does not.

But granted that causes are in general sufficient in the circumstances, in this strong sense, as well as necessary in the circumstances, for their effects, while neither relation holds in non-causal sequences, we can still ask whether in calling something a cause we require both of these features or only one, and if so which.

To clear up this problem, let us consider three different shilling-in-the-slot machines, *K*, *L*, and *M*. Each of them professes to supply bars of chocolate; also each of them has a glass front, so that its internal mechanism is visible. But in other respects the three are different. *K* is deterministic, and conforms to our ordinary expectations about slot-machines. It does not always produce a bar of chocolate when a shilling is put in the slot, but if it does not there is some in principle discoverable fault in or interference with the mechanism. Again, it can be induced to emit a bar of chocolate without a shilling's being inserted, for example by the use of some different object which sufficiently resembles a shilling, or perhaps by poking at the mechanism with pieces of wire. Inserting a shilling is neither absolutely necessary nor absolutely sufficient for the appear-

ance of a bar of chocolate, but in normal circumstances it is both necessary and sufficient for this. ('Necessary' and 'sufficient' are here being used with reference to the machine's laws of working, they describe general relations rather than relations between single occurrences. But it will be a consequence of these general relations that if on a particular occasion, in normal circumstances, a shilling is inserted, a bar of chocolate will come out, and further that the inserting of a shilling on this particular occasion was both necessary in the circumstances and sufficient in the circumstances in the strong sense for this result.) *L*, on the other hand, is an indeterministic machine. It will not, indeed, in normal circumstances produce a bar of chocolate unless a shilling is inserted, but it may fail to produce a bar even when this is done. And such failure is a matter of pure chance. *L*'s failures, unlike *K*'s, are not open to individual explanation even in principle, though they may be open to statistical explanation. With *L*, in normal circumstances, putting a shilling in the slot is necessary, but not sufficient, for the appearance of a bar of chocolate. *M* is another indeterministic machine, but its vagaries are opposite to *L*'s. *M* will, in ordinary circumstances, produce a bar of chocolate whenever a shilling is inserted; but occasionally, for no reason that is discoverable even in principle, the mechanism begins to operate even though nothing has been inserted, and a bar of chocolate comes out. With *M*, in normal circumstances, putting a shilling in the slot is sufficient, but not necessary, for the appearance of a bar of chocolate.

Now on some occasion I put a shilling into *K* and a bar of chocolate comes out. As I have said, putting in the shilling was, in the circumstances, both sufficient in the strong sense and necessary for this result, and we have no hesitation in saying that it caused this result.

Again, I put a shilling into *L* and receive a bar of chocolate. Putting in the shilling was, in the circumstances, necessary for this result. It was also sufficient in the circumstances in the weak sense, but not in the strong, counterfactual, sense. A possible world, with the same laws of working as the actual world, can contain the same circumstances, can lack the result, and yet still contain the inserting of the shilling. The statement, 'Given the circumstances, if the chocolate had not been going

to appear, the shilling would not have been inserted' is not now acceptable. But would we say in this case that the inserting of the shilling caused the appearance of the bar of chocolate? I think we would. Our ordinary causal concept seems to require that where the shilling is put in, the mechanism operates, and a bar of chocolate appears, and would not have appeared if the shilling had not been inserted, the insertion of the shilling caused the appearance of the chocolate despite the fact that in the circumstances even given that the shilling was inserted, the chocolate might not have appeared.

Similarly, I put a shilling into *M* and receive a bar of chocolate. Putting in the shilling was this time sufficient in the circumstances, in the strong sense as well as in the weak sense, for this result. In an appropriate possible world in which the chocolate did not appear, the shilling would not have been put in. But putting in the shilling was not, it seems, necessary in the circumstances. Not only generally but also on this particular occasion the chocolate might have appeared even if no shilling, or anything else, had been put in. But there is room for dispute here. Perhaps, it might be argued, the insertion of the shilling may have been necessary in the circumstances on this particular occasion, though not generally; it may be that on this occasion the mechanism would not have operated if the shilling had not been put in. Moreover, it may be possible to settle this issue. We can, by hypothesis, see the works of the machine. If on this occasion only the later stages of the mechanism operated, not those earlier ones which are normally actuated directly by the shilling, we can decide that the chocolate would have come out anyway, that the insertion of the shilling was not necessary in the circumstances. But if the whole mechanism operated, the issue cannot be settled. Since *M* is, by hypothesis, indeterministic, there is in principle no discoverable answer to the question whether the chocolate would on this occasion have come out if the shilling had not been put in, or, therefore, to the question whether the insertion of the shilling was on this occasion necessary in the circumstances for the result. And yet, it seems to me, it is just this question that we need to have answered before we can say whether the insertion of the shilling caused the result. If the chocolate would not have come out if the shilling had not been put in, then the insertion of the

shilling caused the result. But if it would have come out anyway, the insertion of the shilling did not cause this. (This last ruling prejudices a question about causal over-determination that has still to be considered; but we shall reach an answer to this question which agrees with the present ruling.) And, consequently, if it is in principle undecidable whether the chocolate would on this particular occasion have come out if the shilling had not been put in, it is equally undecidable whether the putting in of the shilling caused the appearance of the chocolate.

The contrast between the comments we have made about the two indeterministic slot-machines *L* and *M* seem to show that 'X caused Y' entails 'X was necessary in the circumstances for Y' and also, trivially, 'X was sufficient in the circumstances for Y' in the weak sense, but *not* the latter in the strong counterfactual sense. This is not required, though in general it holds for what we recognize as causes, at least in the physical sphere. Indeterministic machines are pretty rare.

This conclusion is confirmed by what we are prepared to say about human beings without prejudging the question whether they are wholly deterministic or not. 'Tom's saying what he did caused Bill to hit Tom' is acceptable even if we suppose that Bill had free will, that in the circumstances even after Tom's remark Bill might have controlled himself, provided that we believe that in the circumstances Bill would not have hit Tom if Tom had not said what he did. In this respect we treat Bill in the way I have proposed to treat slot-machine *L*.

There are, however, several possible objections to the thesis that 'X caused Y' entails 'X was necessary in the circumstances for Y'.

The first of these concerns what we may call quantitative over-determination. In sequence A, the hammer-blow caused the chestnut's becoming flatter. But the whole of the blow was not necessary for this result, though it was more than sufficient: a somewhat lighter blow would have sufficed. Even if part of the hammer-head had been absent, this result would still have come about. And so on. But this difficulty is easily overcome. It is possible to go on to a functional dependence view of causation, which would relate exact quantities on the cause side with exact quantities on the effect side. This development will be considered in Chapter 6. But for the present we are dealing only with a fairly primitive causal concept, which

treats events, states, and occurrences, as ordinarily recognized and described, as causes and effects. From this point of view we regard the hammer-blow as a unit, and simply do not consider parts or subdivisions of it or quantitative alterations to it. The alternatives considered are that I strike the chestnut in the way described and that I do not. In constructing possible worlds, in considering what might or would have happened, we either plug in the hammer-blow as a whole or leave it out as a whole. Reducing it a little bit is simply not an option at present under consideration. From this point of view the hammer-blow was necessary in the circumstances: leave it out as a whole from your possible world, and the chestnut remains round.

A much more serious objection concerns what we may call, by contrast, alternative over-determination, or what have been called fail-safe causes. Let us list some examples, some of which are favourites with writers on moral and legal responsibility.

(i) A man is shot dead by a firing squad, at least two bullets entering his heart at once, either of which would have been immediately fatal.

(ii) Lightning strikes a barn in which straw is stored, and a tramp throws a burning cigarette butt into the straw at the same place and at the same time: the straw catches fire.

(iii) '... conditions (perhaps unusual excitement plus constitutional inadequacies) [are] present at 4.0 p.m. that guarantee a stroke at 4.55 p.m. and consequent death at 5.0 p.m.; but an entirely unrelated heart attack at 4.50 p.m. is still correctly called the cause of death, which, as it happens, does occur at 5.0 p.m.'

(iv) Smith and Jones commit a crime, but if they had not done so the head of the criminal organization would have sent other members to perform it in their stead, and so it would have been committed anyway.

(v) A man sets out on a trip across the desert. He has two enemies. One of them puts a deadly poison in his reserve can of drinking water. The other (not knowing this) makes a hole in the bottom of the can. The poisoned water all leaks out before the traveller needs to resort to this reserve can; the traveller dies of thirst.<sup>5</sup>

<sup>5</sup> I touched on this problem in 'Causes and Conditions', *American Philosophical Quarterly*, ii (1965), 245-64, esp. 250-2. Examples (ii) and (iv) are borrowed from



In each of these five cases, we cannot say of either of the candidates for the role of cause that it was necessary in the circumstances for the effect. If either of the two bullets had not been fired, the man would still have died. If the lightning had not struck, the straw would still have caught fire; and equally if the tramp had not thrown the butt away. If Smith and Jones had suddenly abandoned a life of crime, the same crime would still have been committed by the other members of the gang; but equally if those other members had not been ready to act, the crime would still have been committed, as it was, by Smith and Jones. If the heart attack had not occurred, the stroke would have carried the man off at 5.0 p.m.; but if the conditions for the stroke had been absent, the heart attack would still have killed him as and when it did. If the can had not been punctured, the traveller would have died of poison, perhaps even sooner than he actually died of thirst; but if it had not been poisoned, he would have died just as and when he did.

But though they have this common feature, these examples fall into two groups. In (iii) and (iv) we have no hesitation in making specific causal statements, that the heart attack caused the man's death, and that the actions of Smith and Jones brought about the criminal result. In (v) also it seems clear to me that the puncturing of the can caused the traveller's death, but as Hart and Honoré say that 'it is impossible to give a satisfactory answer to this question in terms of either *B* or *C* or both causing *A*'s death', and that 'their mutual frustration of each other's plan precludes us from saying that either caused *A*'s death', I need to explain and defend my view of this case.

Where we have no hesitation in making causal statements we can tell some more detailed causal story: we can say how the heart attack caused the man's death, how Smith and Jones committed the crime. But the rival story about the alternative or reserve cause cannot be completed. The conditions for the stroke did not actually lead to a stroke, and since there was no stroke, no stroke led to death. The reserve members of the

K. Marc-Wogau, 'On Historical Explanation', *Theoria*, xxviii (1962), 213-33, the latter coming originally from P. Gardiner, *The Nature of Historical Explanation*, p. 101. Example (iii) is quoted from M. Scriven, review of E. Nagel, *The Structure of Science*, in *Review of Metaphysics* (1964). Example (v) is based on a modification by Hart and Honoré (*Causation in the Law*, pp. 219-20) of a hypothetical case devised by J. A. McLaughlin, *Harvard Law Review*, xxxix (1925-6), 149, 155 n. 25.

gang remained in reserve. What we accept as causing each result, though not necessary in the circumstances for that result described in some broad way, was necessary in the circumstances for the result *as it came about*.

This matter can be thoroughly clarified if we introduce here a distinction which will be discussed more fully in Chapter 10, between *facts* and *events* both as causes and as results or effects. In (v) the puncturing of the can brought it about that the traveller died of thirst, that is, it caused his dying of thirst (though it prevented his dying of poison). But we cannot say that the puncturing of the can brought it about that he died, or caused his dying—since he would have died anyway, if it had not been punctured. *That he died*, and *that he died of thirst*, are distinguishable facts, and hence distinguishable results. So, as long as we are dealing with fact-results, it is not surprising that the puncturing of the can should have brought about the second of these but not the first. But if we think of an effect as a concrete event, then the event which was the traveller's death was also his death from thirst, and we must say that the puncturing of the can caused it, while the poisoning did not. For a concrete event effect, we require a cause, or causal chain, that leads to it, and it is the chain puncturing-lack-of-water-thirst-death that was realized, whereas the rival chain that starts with poison-in-can was not completed. In this way I would defend my judgement that the traveller's death—that is, the concrete event—was caused by the puncturing of the can, but at the same time explain the doubts felt by lawyers (and others) about this case as due to the equally correct judgement that his dying—that is, the fact that he died on this journey—was not caused by the puncturing of the can.

These distinctions vindicate our general thesis that '*X* caused *T*' entails '*X* was necessary in the circumstances for *T*': apparently conflicting answers arise from some uncertainty or equivocation about what is to count as *T*, whether it is a concrete event or a fact, and if so, which fact. If *T* is the fact that the traveller died of thirst, then the puncturing of the can both caused and was necessary in the circumstances for *T*. If *T* is the fact that the traveller died on this journey, then the puncturing of the can neither caused *T* nor was necessary in the circumstances for *T*. If *T* is the traveller's death qua concrete

event, which was, among other things, a death from thirst, the puncturing of the can both caused and was necessary in the circumstances for  $T$ .

In the dubious cases of alternative over-determination, such as (i) and (ii), it is natural to reject such statements as 'This bullet caused his death' and 'The lightning caused the fire'. In these cases even a detailed causal story fails to discriminate between the rival candidates for the role of cause, we cannot say that one rather than the other was necessary in the circumstances even for the effect *as it came about*. Even if we take the effect as a concrete event all that was clearly necessary for it in the circumstances was in (i) the volley and in (ii) the lightning/cigarette-butt cluster of events, and it is such clusters that we can confidently take as causing these effects. 'But which item in the cluster really caused (or "brought about") this effect (or "result")?' is a sensible question in so far as it asks for a discrimination between the alternative 'causes' by way of the filling in of a more detailed account. But if no more detailed correct account would provide the desired discrimination, this question has no answer.

For instance, if, as in an example mentioned earlier, a match is struck and touched with a red-hot poker at the same time, and a flame appears, we can say that the striking and the touching together caused the flame to appear, since if neither of them had occurred in the circumstances no flame would have appeared, but if the match was affected in both these ways simultaneously, we cannot say that either by itself caused this. Our ruling in these cases agrees with what we said about slot-machine  $M$ , where a shilling was put in and a bar of chocolate came out, but the bar might have come out even if no shilling had been put in: if on this particular occasion a bar of chocolate *would* (not merely *might*) have come out anyway, then the insertion of the shilling did not cause its appearance. This is a strange case of causal over-determination, because one of the rival factors is a null one: the bar would, on the present hypothesis, have come out spontaneously. But even a null alternative in a case of alternative over-determination is enough to prevent the other alternative from causing the result.

So far, then, we can defend the conclusion reached earlier: the statement that  $X$  caused  $T$  entails that  $X$  was necessary in

the circumstances for  $X$ , but it does not entail that  $X$  was in the strong sense sufficient in the circumstances for  $X$ . But the latter often holds as well.

Something very like this is the main negative thesis of Professor Anscombe's Inaugural Lecture: contrary to a long dominant philosophical tradition, causes need not be 'sufficient' for their effects, need not 'necessitate' them. Raymond Martin also has defended a 'necessity thesis' while criticizing a 'sufficiency thesis' which he finds in an earlier work of mine.<sup>6</sup>

But is there anything still to be said for the sufficiency thesis? A corollary of its rejection is that causal consequence is *not* formally analogous to logical consequence, legal consequence, and so on. If  $Q$  is a logical consequence of  $P$ , then  $P$  is logically sufficient for  $Q$ . If, given  $R$ ,  $Q$  is a logical consequence of  $P$ , then, given  $R$ ,  $P$  is logically sufficient for  $Q$ . Similarly if  $Q$  is a legal consequence of  $P$ , then  $P$  is, in the circumstances, legally sufficient for  $Q$ . And this holds in general for other sorts of consequences. But by our present account,  $Q$  may be a causal consequence of  $P$  without  $P$  being causally sufficient for  $Q$  even in the circumstances. This is strange, and yet it seems correct. There being a radon atom here now is a causal consequence of there having been a radium atom here a little earlier, but if radioactive decay is a non-deterministic process there are no circumstances such that in them the earlier presence of the radium atom was sufficient (in the strong, counterfactual sense) for the radon atom's being here now: the former might not have decayed, and even if it had the latter might also have decayed already. Similarly if Jim, who is colour-blind, and his wife Alice, who carries no gene for colour-blindness, have a daughter, Jane, who has a son, Tom, and Tom is colour-blind, then Tom's being colour-blind is a causal consequence of Jim's being colour-blind; but the latter gave only a fifty-fifty chance that any son of Jane would be colour-blind.

On the other hand, if  $P$  caused  $Q$ , we can surely say that  $Q$  occurred because  $P$  did; the latter is practically equivalent to 'Since  $P$  occurred,  $Q$  occurred'; a since-statement can be fairly

<sup>6</sup> G. E. M. Anscombe, *Causality and Determination*; Raymond Martin, 'The Sufficiency Thesis', *Philosophical Studies*, xxiii (1972), 205-11; Martin, who criticizes my 'Causes and Conditions', *American Philosophical Quarterly*, ii (1965), 245-64, sent me an earlier version of his paper in 1969.

aptly described, in Goodman's terms, as a factual conditional, that is, as adding to the open non-material conditional 'If *P* occurred, *Q* occurred' the presupposition that *P* did occur (and hence that *Q* did so too). Putting these steps together, we seem compelled to take '*P* caused *Q*' as entailing the non-material conditional 'If *P* occurred, *Q* occurred', which would commit its user to the counterfactual 'If *Q* had not been going to occur, *P* would not have occurred', that is to the strong counterfactual sufficiency of *P* for *Q*.

Another argument on the same side is that a future causal statement '*P* will cause *Q*' or '*P* would cause *Q*' seems to claim that *P* is sufficient in the circumstances for *Q*, that if *P* occurs, *Q* will occur also, and that if *Q* is not going to occur, neither is *P*. Even if we were right in saying that a past tense singular causal statement entails only that the cause was necessary, not also sufficient, in the circumstances for the effect, this may be characteristic of past tense uses rather than of causal statements in general.

It seems, then, that the analysis of a tenseless singular causal statement '*P* causes *Q*' should include both the non-material conditionals 'If *P* occurs, *Q* occurs' and 'If *P* does not occur, *Q* does not occur', and therefore also their contrapositives, each of which will change into a counterfactual if the speaker takes its antecedent not to be fulfilled. But different parts of this concept will be stressed in different settings. When *P* and *Q* are both known to have occurred, it is natural to lay stress on the counterfactual form of the second conditional; but if their occurrence is problematic, it is natural to emphasize the open form of the first conditional, though the second is also in force. The general notion of a cause is of something which is both necessary and sufficient in the circumstances for its effect, but where the cause and the effect have both actually occurred we do not require that the cause should be sufficient in the strong counterfactual sense.

Some writers who have recognized that causes need not be sufficient for their effects have drawn the further conclusion that our ordinary concept is primarily probabilistic. Thus Patrick Suppes<sup>7</sup> argues that 'the everyday concept of causality is not sharply deterministic' from the evidence that we may say 'His

<sup>7</sup> *A Probabilistic Theory of Causality*, p. 7.

reckless driving is bound to lead to an accident', where '“lead to” conveys the causal relation between the reckless driving and the predicted accident' and yet '“is bound to” means [only] that the probability is high'. But this is a confusion. Saying that *A* is likely to cause *B* does not put likelihood into the causing itself: it could (though it does not) mean that *A* is likely to necessitate *B*, and I think it does mean that *A* is likely to be, in some particular case, necessary in the circumstances for *B*. What will fulfil this probabilistic prediction is an actual crash for which the man's reckless driving was necessary in the circumstances. Similarly Suppes is right in claiming that a mother who says that her child is afraid of thunder does not mean that a state of fright ensues whenever the child hears thunder, but rather that there is a fairly high probability of this; but when the child is actually frightened by thunder, the thunder is necessary in the circumstances for its state of fright.

Thus it may often happen that the only fully explicit causal generalization that a speaker is prepared to make is a probabilistic one, and yet he may still be taking it that the causal relation in any particular case that fulfils it is one of necessity, and perhaps also sufficiency, in the circumstances. To go straight from this evidence to a probabilistic theory is a mistake, and one that results from holding on to the Humean doctrine that causation is essentially general while admitting that causal claims need not involve *universal* generalizations. Against this I shall argue (in Chapter 3) for the primacy of singular as opposed to general causal statements.

These points seriously undermine Suppes's project of analysing causation as a whole in terms of probabilities. We can admit, none the less, that there may be probabilistic or statistical laws, and singular sequences of events which in some sense fulfil them, and that both of these can be called causal in a broad sense. These will be discussed in Chapter 9, with particular reference to problems about the form and content of statistical laws and the interpretation of the 'probabilities' associated with them—problems which Suppes does not bring out or resolve. For the present, however, we can affirm that a cause is ordinarily taken to be necessary in the circumstances for its effect, and perhaps sufficient in the circumstances as well.

We cannot, however, conclude that necessity-in-the-circum-

stances is *the* distinguishing feature of a cause. As we saw, one collateral effect may be necessary in the circumstances for another, but it is not said to cause the other. Also, if  $X$  caused  $Y$ , we have seen that  $X$  may be in the strong sense sufficient in the circumstances for  $Y$ . If it is, then  $Y$  will be necessary in the circumstances for  $X$ ; but we shall not say that  $Y$  caused  $X$ .

Such counter-examples show that another relation, which we may call *causal priority*, is thought of as being required, along with necessity-in-the-circumstances, for causing. Of a pair of collateral effects, neither is causally prior to the other, and in our other example, since  $X$  caused  $Y$ ,  $X$  was causally prior to  $Y$ , and  $Y$  was therefore not causally prior to  $X$ . But it seems that if any  $X$  is both necessary in the circumstances for and causally prior to  $Y$ , we shall say that  $X$  caused  $Y$ ; also, wherever we are prepared to say that  $X$  caused  $Y$  we are prepared to say that  $X$  was necessary in the circumstances for and causally prior to  $Y$ . The distinguishing feature of causal sequence is the conjunction of necessity-in-the-circumstances with causal priority.

But this conclusion is rather hollow until we have said more about causal priority. A full discussion of this notion, and of the whole problem of the direction of causation, will be undertaken in Chapter 7; but a preliminary account will be in place here.

The core of the notion of causal priority is that the world has some way of running on from one change to another. But we can speak with any accuracy only when we associate causal priority with some kind of sufficiency or necessity. Thus  $X$  is sufficient in the circumstances in the weak sense for and causally prior to  $Y$  provided that if  $X$  is put into the world in the circumstances referred to and the world runs on from there,  $Y$  will occur. Similarly  $X$  is necessary in the circumstances for and causally prior to  $Y$  provided that if  $X$  were kept out of the world in the circumstances referred to and the world ran on from there,  $Y$  would not occur. And  $X$  is sufficient in the circumstances in the strong sense for and causally prior to  $Y$  provided that if  $X$  were put into the world in the circumstances referred to and the world ran on from there,  $Y$  would occur. Sufficiency-in-the-weak-sense plus causal priority is shown by an experiment on the actual world; but necessity and sufficiency-in-the-strong-sense plus causal priority involve counterfactual claims, and

therefore involve assertions about how the world would have run on if something different had been done: they involve thought about the independent running of a merely possible world.

We have already illustrated this in suggesting a way out of the difficulty about collateral effects. We thought of constructing a possible world which was just like the actual world up to the point where, in the actual world, the proposed antecedent occurs, of excluding this antecedent from the possible world and then letting that world run on: if and only if the result did not occur in that possible world were we prepared to say that the antecedent was *in the special sense* necessary in the circumstances for the result. That 'special sense' of 'necessary in the circumstances' is what we now recognize as the conjunction 'necessary in the circumstances for and causally prior to'.

But this account needs some adjustment to make it generally applicable. Our proposal to make the possible world just like the actual one up to the point of time at which the proposed antecedent  $X$  occurs in the actual world would have the undesired result of making temporal priority necessary, though not sufficient, for causal priority. For if  $Y$  had occurred before  $X$  in the actual world, it would already have been included, by the above proposal, in the possible world: we should have left no chance that the exclusion of  $X$  from the possible world might, as that world ran on, exclude  $Y$  also. But temporal priority is not conceptually necessary for causal priority. We can coherently consider the possibility of backwards or time-reversed causation. To leave room for this, we must modify the suggested account. We must think of somehow excluding an  $X$  from a possible world at a time  $t_3$  (where  $X$  occurred in the actual world at  $t_3$ ) without prejudice to the question whether a  $Y$  occurred in that possible world (as it did in the actual one) at an earlier time  $t_2$ . But to combine this with the notion of letting the possible world run on, we must think of excluding  $X$  indirectly, by excluding some other possible event  $W$  which actually occurred at a still earlier time  $t_1$  and which is believed to be necessary in the circumstances for  $X$ . Our possible world is to have the same *circumstances* as the actual world, but to diverge from it by the exclusion of  $W$  at  $t_1$ ; from there on it is allowed to run on by whatever laws of working the actual



world obeys. This leaves open the possibility that there should be among these laws a backward causal one such that this indirect exclusion of  $X$  would carry with it the non-occurrence of  $Y$ . If  $Y$  does occur in our possible world at  $t_2$ , then  $X$  is not necessary in the circumstances for and causally prior to it; but if  $Y$  does not occur in our possible world at  $t_2$ ,  $X$  may be so, though there are other possibilities which it is not easy to exclude. (These details are discussed in Chapter 7.)

The notion of necessity-in-the-circumstances-plus-causal-priority, then, involves the thought of what would happen when a possible world, constructed by some appropriate alteration from the actual world, was allowed to run on. It presupposes that the actual world has some laws of working—not necessarily strictly deterministic ones—which can be carried over to the possible world. But it does not require that in order to use this notion we should know what those laws of working are.

I have explicitly added the notion of causal priority to that of being necessary in the circumstances. But I must admit that the terms 'necessary' and 'sufficient' are often so used as to include a suggestion of causal priority: even where a cause is sufficient in the circumstances in the strong sense for its effect we find it strange to say that the effect is necessary in the circumstances for the cause, just because the effect is not causally prior to the cause. However, there are senses of 'necessary' and 'sufficient' which do not include causal priority and which are exhausted by the appropriate conditional analyses, and I think it makes things clearer if we use 'necessary' and 'sufficient' in these simpler senses, and introduce causal priority as a further element.

In so far as we have used conditionals, especially counterfactual ones, in our analysis of the concept of causation, that analysis is incomplete until these conditionals have themselves been explained. A view of conditionals has probably been implicit in what I have already said; to make it explicit I want simply to take over conclusions that I have reached elsewhere.<sup>8</sup> A non-material conditional statement introduces a supposition (the antecedent) and asserts something (the consequent) within the scope of that supposition. The conditional 'If  $P$ ,  $Q$ ' can be paraphrased by 'Suppose that  $P$ ; on that supposition,  $Q$ '; or

<sup>8</sup> In Chapter 3 of *Truth, Probability, and Paradox*.

again by 'In the possible situation that  $P$ ,  $Q$  also'. This account holds for all non-material conditionals; a counterfactual adds to this the suggestion that the antecedent does not hold in the actual world. If so, there is liable to be a certain arbitrariness in the choice of 'the possible situation that  $P$ ', or in other words of the particular possible world of which the consequent is being asserted. It is a consequence of this analysis that only some conditionals are capable of being true or false. A conditional whose antecedent is fulfilled will be true if its consequent is true and false if its consequent is false. But if the antecedent of a counterfactual is, as its user believes, unfulfilled, it cannot be true or false in this way. A counterfactual whose antecedent entails its consequent can be allowed to count as true. But most counterfactuals cannot achieve truth in this way either, and will be neither true nor false, though they may be acceptable or unacceptable, well or poorly supported, and so on.

We are led, then, towards two conclusions which are similar to Hume's. Statements of singular causal sequence involve in their analysis counterfactual conditionals which on the present showing are not capable of being true; so the singular causal statements cannot be true either.<sup>9</sup> (They can still be false in rather obvious ways: ' $X$  caused  $Y$ ' will be false, for example, if either  $X$  or  $Y$  did not occur.) Also, whether they can be true or not, these counterfactual conditionals describe possible situations or possible worlds, they are concerned with suppositions and their consequences or accompaniments, they do not describe what actually occurred, let alone what was observed to occur, in the actual individual sequence. They state what would have happened, not what did happen. Necessity<sub>1</sub>, then, the distinguishing feature of causal sequences, is not something that can be observed in any of those sequences. The case for saying this is even stronger than Hume's case for saying the same about necessity<sub>2</sub>. Consequently a question analogous to one of Hume's arises here: how do we acquire the idea of necessity<sub>1</sub>, since we do not derive it from anything we observe in the individual causal sequences to which we apply it? It looks as if a psychological account is called for; as I said at the

<sup>9</sup> A similar conclusion is reached by A. J. Ayer, *Probability and Evidence*, pp. 132-9. But for me this is only a tentative conclusion, which will be qualified at the end of Chapter 8.

end of Chapter 1, Hume's resort to psychology was not wrong in principle but merely premature.

The item for which we must seek a psychological explanation is not, as Hume thought, the idea of a support for *a priori* inferences from one distinct existence to another, but the sort of thinking that is expressed by the counterfactual conditional, 'If in these circumstances *X* had not occurred, *Y* would not have occurred either', coupled with the notion of *X*'s causal priority to *Y*, the thought that the actual world ran on from *X* to *Y*, and that the appropriate possible world would run on from the exclusion of *X* to the non-occurrence of *Y*. Let us take the counterfactual first, leaving aside the priority. It can be paraphrased as 'Suppose that in these circumstances *X* did not occur; then *Y* did not occur either' or 'Given these circumstances, in the possible situation that *X* did not occur, *Y* did not occur'. For the general capacity and tendency to make suppositions, to consider possibilities that appear not to be realized, I can offer no explanation beyond the obvious evolutionary one; this capacity and tendency are of some value to their possessors and could have been fostered by the ordinary processes of natural selection. But it is worth stressing here that although we express this kind of thinking in words, its development need not have waited for the development of language. It is not essentially tied to verbal expression, and there is some reason to suppose that other, non-verbal, animals share this capacity with us. Possibilities can be literally envisaged rather than described.<sup>10</sup> But, given that we have this general tendency, can we explain the particular ways in which we employ it? That is, can we explain how, having supposed *X* not to have occurred in these circumstances, we decide what else to assert within the scope of this supposition, or how to fill out our picture of this merely possible situation? We can distinguish a sophisticated and a primitive way of doing this. The sophisticated way uses general propositions which we take to be confirmed by observations of the actual world, but which we feel justified in extending beyond the instances in which they are confirmed not only to other actual instances but to merely possible ones which are related to the confirming instances in

<sup>10</sup> Cf. *Truth, Probability, and Paradox*, p. 100.

the same way that other actual instances would be.<sup>11</sup> These are combined with the supposition, and consequences drawn by deductive reasoning from the combination are asserted within the scope of the supposition. This sophisticated way of developing a supposition anticipates a well-known procedure of 'natural deduction'. The primitive way of doing the same job relies not on this combination of inductive and deductive reasoning, but on imagination and analogy. I have observed another situation very like the present one, in which (unlike the present one) no event of the *X* type occurred. I borrow features from that other situation to fill out my imaginative picture of the possible situation that in the present circumstances *X* did not occur. In particular, if no event of the *Y* type occurred in that other situation, I see the possible situation that *X* did not occur in the present circumstances as continuing without the occurrence of *Y*. The sort of observation that can be immediately used in this imaginative analogizing is that prescribed for Mill's Method of Difference: two otherwise very similar instances, in one of which both the 'cause' and the 'effect' occur, and in the other of which neither of these occurs. Such a pair of instances may be found in a number of different ways, but the most obvious, and presumably the most fruitful original source of the causal concept, is provided by the 'before and after' observation. In an otherwise apparently static situation, one striking change (*X*) occurs, followed shortly afterwards by another (*Y*). The situation before *X* occurred, when equally *Y* did not occur, provides the control case or negative instance, while the later situation, in which both *X* and *Y* occurred, provides the experimental case or positive instance. And in a quite primitive and unsophisticated way we can transfer the non-occurrence of *Y* from the before situation to a *supposed* later situation in which, similarly, *X* did not occur, and form the thought which is expressed by the statement 'If *X* had not occurred, *Y* would not have occurred', or, in other words, '*X* was necessary in the circumstances for *Y*'.

But this is only one part of our ordinary idea of necessary connection. We have also to account for the other element, the causal priority. It seems undeniable that this notion arises from our experience of our own active interventions in the world. If

<sup>11</sup> Cf. *Truth, Probability, and Paradox*, pp. 117-18, and Chapter 8 below.

I *introduce* the change  $X$  into an apparently static situation and then the other change  $Y$  occurs, I not only see  $X$  as in the weak sense sufficient in the circumstances for  $Y$ , but also see  $X$  as causally prior to  $Y$ . I see the actual world as running on from my introduction of  $X$  to the occurrence of  $Y$ . If, further, I think (for the reasons just outlined) of  $X$  as necessary in the circumstances for  $Y$ , I see the *possible* world in which I do *not* introduce  $X$  as running on without  $Y$ 's occurring. I seem then to have introduced into the actual world not only  $X$ , but the whole  $X$ - $Y$  sequence; but I seem to have done so by operating on the  $X$  end of it. The notion of the actual world's running on *from*—not *to*— $X$ , and of a possible world's similarly running on from not- $X$ , is derived from this complex experience of intervening and then waiting for results. Rightly or wrongly, I see myself as a free agent in introducing  $X$ , and therefore rule out the possibility that anything else in the objective situation is causally prior to  $X$ , that the world ran on to  $X$ ; I therefore rule out also both the possibility that  $X$  and  $Y$  should be collateral effects of some common cause and the possibility that  $Y$  should have caused  $X$ , even if I see  $Y$ , as I well may, as being necessary in the circumstances for  $X$ .

We can, then, find natural and presumably instinctive ways of thinking which would be expressed by the counterfactual conditionals and assertions of causal priority which we have given as the analysis of necessity<sub>1</sub>, of the distinguishing feature of causal sequence. But just as our analysis of necessity<sub>1</sub> differs from Hume's equation of it with necessity<sub>2</sub>, so our account of the psychological mechanism involved is different. The key item is a picture of what *would* have happened if things had been otherwise, and this is borrowed from some experience where things *were* otherwise. It is a contrast case rather than the repetition of like instances that contributes most to our primitive concept of causation.

My main concern in this chapter has been to analyse our actual concept of causing, to formulate what we commonly take to be the distinguishing mark of causal sequences, and then, since this has turned out not to be an observed feature of those sequences, to sketch an explanation of our thinking in this way. In this analysis the major item is that a cause is thought to be both necessary and sufficient in the circumstances for its effect,

but that the sufficiency is less firmly required than the necessity, particularly where the sequence is known to have occurred. But whatever our actual concept may be, it is obvious that we can construct causal concepts as we wish, which may be useful for particular purposes. We can speak, if we like, of *necessary causes*, and make it explicit that sufficiency is not required. Or we can speak of *sufficient causes*, and make it explicit that necessity is not required. Or we can make it plain that we are requiring both necessity and sufficiency in the circumstances. But the general pattern on which such concepts are constructed is that we combine the appropriate non-material conditional(s) with the notion of causal priority. But all such causal concepts are somewhat superficial: there may well be further backing for these conditional descriptions, further relations in the objects that encourage us to speak and think in these ways.