C4 **4**

CAUSALITY

C4.S1 ## 4.1 INTRODUCTION

C4.P1 Things in the world change all the time. These changes (or most of them, anyway[1]) do not just happen by themselves, but because they are *caused* to occur. Since the beginning of rational thought, scientists and philosophers alike have considered it as one of their main objectives to discover and understand the causes of changes in nature. Causation has therefore played a crucial role in nearly every attempt to make sense of the fabric of the world. Immanuel Kant, for instance, thought that causation was one of the necessary categories of thought by means of which we structure the different sensory impressions we have of the world in order to bring them into the order of one single and unified experience. Almost two hundred years later, John Mackie captured the central importance of causation for our understanding of the world in his famous slogan that causation is the "cement of the universe" (1974); for causation, according to Mackie, is what holds together earlier and later stages in the

C4.N1 1. Some people would argue that there are also changes which happen spontaneously: just on their own. But we can set the question aside here how such "uncaused changes" would work.

history of the universe, which would otherwise occur as unconnected bits and pieces, just one after the other. However, while nearly all philosophers agree on the crucial importance of causation, there is no consensus at all on how causation itself should be understood: what is it for *x* to cause *y*? And just to give you a brief glimpse of some of the many hotly debated subquestions over which the debate has been raging: is causation an irreducible primitive connection, or can it be (explanatorily, or metaphysically) reduced to something else? What kind of entities stand in causal relations to one another: only events, or also substances? Or facts, rather than events? Only particular entities, or abstract entities as well? How is causation related to explanation? And is causation an extensional relation, which holds between two entities no matter how these entities are described, or an intentional one, which holds between entities only when they are described in particular ways?

C4.P2 Most of the modern debate about causation has been shaped by the influential empiricist tradition whose main historical proponent has been David Hume, and which has attempted to reduce causation to the obtaining of (certain types of) regularities. As we have already seen in chapter 2 (section 2.4), Hume criticized the concept of causal power on the grounds that it failed to meet the empiricist criterion for meaningfulness which he advocated. This criterion required that we could trace back concepts to original impressions (as Hume puts it: all our ideas "are copied from impressions").[2] Inherently bound up with Hume's critique of the concept of power was his critique of the "traditional"

C4.N2 2. See *A Treatise of Human Nature*, 4, and *An Enquiry Concerning Human Understanding*, 9.

concept of causation. This is not surprising, because Hume considered the notion of "necessary connection," which on the traditional view was part and parcel of the notion of cause, as practically synonymous to the notion of power.[3] The view that a cause is somehow necessarily connected to its effect—the cause "makes" its effect occur, and the effect "cannot help occurring" once the cause is active—is both intuitively appealing and has been historically highly influential. Causes, we normally think, make things happen, and this is what allows us to *explain* these effects by referring to their causes. But, Hume argued, we have no "original experience" of necessitation that could serve as the original experience to which the concept of causation could be traced back. Focusing on presumed causal connection between events (i.e., cases where presumably one event causes another one), Hume wrote:

C4.P3 > There is not, in any single, particular instance of cause and effect, anything which can suggest the idea of power or necessary connection. (*An Enquiry Concerning Human Understanding*, 63)

C4.P4 When we consider events or objects themselves, and consider particular (supposed) instances of causation, Hume thought, we would never find any real element corresponding to the idea of a causal nexus or a necessary connection. What we do observe is only that certain types of events are generally conjoined: events of one type generally follow upon events of another type.

C4.N3 3. See *A Treatise of Human Nature*, 157.

C4.P5 > All events seem entirely loose and separate . . . one event follows another; but we can never observe any tie between them. They seem *conjoined*, but never *connected*. (*An Enquiry Concerning Human Understanding*, 74)

C4.P6 Having observed this general conjunction, we come to expect, when we see an event of the first type, that an event of the second type will follow. And this habitual conjunction we then project onto the world, thinking that the connection we make in our mind mirrors something *really out there* in the succession of events itself. The source of our idea of causation is therefore, on Hume's account, something merely subjective.

C4.P7 The projectivist or antirealist element of Hume's own view—which claims that causation was "nothing real in the world" but merely a projection of our minds—was, however, not to be as influential as the positive thesis about causation that could be developed from Hume's arguments, namely, that causation could be analyzed in terms of two other ideas: (i) temporal succession and (ii) constant conjunction. Take Hume's own proposed "definition" of "cause" in the *Enquiry*:

C4.P8 > A cause is an object, followed by another, where all objects similar to the first are followed by objects similar to the second. (*An Enquiry Concerning Human Understanding*, 76)

C4.P9 Or, to put this more precisely: for x to be a cause of y, (i) y has to follow x in time and (ii) there must be a constant and regular pattern of items like y following upon items like x. This proposal has three important implications for the nature of causation, which are all far from uncontroversial, but

which have become key tenets of the Humean "orthodoxy" on causation.

C4.P10 First, cause and effect must be things that can "follow one another," and, despite Hume's talk of "objects" this best fits the case where one *event* causes another. By contrast, substances, which most of the philosophical tradition prior to Hume had considered to be paradigmatic causes, can rarely be said to "follow one another"—except in the special sense in which descendants can be said to "follow" their forebears, which is certainly not a model for standard causal interactions. Nor is it the case in normal causal interactions that an event temporally "follows" a substance: this would require the substance to go out of existence before or when the event occurs, which is, again, a rather special case. As a result, Humeans generally came to hold that the causal relata are events rather than substances.

C4.P11 Second, on the Humean view, particular instances of causation necessarily presuppose general regularities: for example, your turning the light switch on now can cause the light to go on immediately afterward only if there is some general regularity in the background. This regularity need not directly connect the events we actually perceive (perhaps you have only turned this particular light switch once in your life), but may only connect the "hidden" intermediate steps between visible cause and effect. But still, on this view, there can be no genuinely singularist causation, in the sense that one individual event might cause another one *no matter what else is going on in the world*. Third, on Hume's view we can provide an analysis of causation in terms of things that are not themselves causal. Causal facts are completely determined by the non-causal facts in the world (i.e., the facts about temporal succession and regularities). Once the latter

facts are all fixed, we also know what causes what, and we do not have to introduce the causal facts as independent extras.

C4.P12 Contemporary Humeans still mostly accept these three claims, even though they disagree with Hume with respect to the kind of general pattern that supposedly underlies individual instances of causation. Most contemporary neo-Humeans believe that such general patterns are not just regularities that hold de facto, but must be genuine laws of nature that support counterfactual statements. For this reason, their views are also called "nomological" views of causation. As they have shaped so much of the modern debate, nomological views of causation that follow Hume in rejecting singular causation, and in taking causal facts to be determined by non-causal facts, are a useful point of departure for our following discussion. In the next section (4.2), we will consider two influential recent versions of the Humean approach, and examine some of the problems these accounts encounter. Such problems indicate that there are still major unresolved issues in the Humean approach. This motivates us to turn to Aristotle's account of causation (section 4.3), which will provide a useful foil for investigating how the shift toward realism with respect to powers that we have discussed in chapter 2 gives us new resources to account for causation in terms of the exercise of causal powers (section 4.4). In that context, we will also consider whether an account of causation in terms in powers can explain the direction of causation, by means of the distinction between active and passive powers. We will conclude the chapter by turning to mental causation, that is, the question of how mental occurrences or entities can cause physical changes and vice versa (section 4.5).

C4.S2

4.2 SOME NEO-HUMEAN DEVELOPMENTS

C4.P13

As we have just seen, contemporary Humeans tend to assume that the regularities that underlie particular causal connections are not just general regularities that hold as a matter of fact, but laws of nature. In making this move, Humeans do not usually take themselves to necessarily be in opposition to Hume himself, who writes, directly after the definition of "cause" in the *Enquiry* that we have quoted earlier: "Or, in other words, *where the first object had not been, the second never had existed*" (*An Enquiry concerning Human Understanding*, 76). These lines suggest that Hume thought of causation not simply as based on general regularities, but (also) in terms of necessary conditions.[4] The latter is a stronger conception, since some regularities are merely contingent. You can imagine, for instance, that, de facto, all men in a certain community put on their hats when going out, but this does not allow you to conclude that if Jim, a man from that community, had not put on his hat, he wouldn't have gone out, because Jim might have had to go out so urgently that he would have done so even if he had forgotten his hat at home.

AQ: In FN 4, single quotes O.K. here, or change to double? Please clarify

C4.P14

Many metaphysicians of neo-Humean allegiance with respect to causation endorse this stronger conception, and have taken laws of nature to be the obvious candidates for

C4.N4

4. At the same time, Hume's emphatic critique of the supposedly 'necessary connection' involved in causation more strongly suggests that he did want to settle for merely regularities. As a consequence, contemporary Humeans, while being generally inspired by him, can hardly claim to defend his own (historical) position.

what explains why a cause is necessary for its effect to occur. If there is a law of nature that an event of type A follows upon an event of type B, then we do have a good explanation of why, ceteris paribus, if the B-type event had not occurred, the A-type event would not have occurred either. As a result, many Humeans, from the nineteenth century onward, came to think of causes as necessary—or, going beyond Hume, necessary *and sufficient*—conditions for their effects (most notably John Stuart Mill). But this had the unwelcome effect of putting the philosophical analysis of cause very much at odds with our ordinary way of talking, both in everyday life and in scientific practice. For we call many factors "causes" that are not, strictly speaking necessary—or sufficient—for their effect. When we regard, for example, the throwing of a stone as a genuine cause of the breaking of the window, the throwing of the stone was neither necessary for the latter (since the window could have broken in many other ways as well) nor sufficient (since the window would not have broken had something arrested the stone in its flight). Presumably, in most cases, only very encompassing states of affairs concerning the whole environment of a change are truly sufficient, or necessary, for bringing about the effect, but we are prepared to call less encompassing factors "causes," too.

C4.P15 One of the most influential Humean attempts to resolve this discrepancy is due to J. L. Mackie, who in his *The Cement of the Universe: A Study of Causation* (1974) tried to analyze the relation between particular causes and particular effects in terms of a sophisticated combination of necessary and sufficient conditions. As Mackie argued, a cause is not (normally) sufficient for its effects, nor is it necessary. To take Mackie's own example: a short circuit in a house causes a fire. The short circuit is not sufficient on its own to cause the fire,

because without the presence of inflammable material and oxygen (etc.), the short circuit would not have brought about the fire. Neither is the short circuit necessary, since something other than the short circuit could, in principle, have caused the fire as well. Instead, Mackie claims, the short circuit is an **I**nsufficient but **N**ecessary part of a set of conditions which is itself **U**nnecessary but **S**ufficient (for the effect). The cause is thus what he calls an **INUS** condition. The (overall) sufficient-condition is a very large collection of contributory conditions that absolutely suffices for the effect on this occasion. Without the short circuit this collection would not have sufficed for the effect, though, and would not have led to the fire. Mackie's analysis allows us to call factors "causes" that are by themselves neither necessary nor sufficient for the effects. At the same time, it captures the idea that in some respect, namely under the very precise conditions that obtained, the cause *was* necessary and sufficient for the effect. For given the other conditions, it was both. Thus, given the presence of the oxygen and of the inflammable material as well as the absence of possible interveners that would have prevented the outbreak of the fire, the short circuit *was* sufficient for the outbreak fire. At the same time, given these other conditions and given that there was no alternative factor that would have led to the fire in combination with them, the short circuit was still necessary for these other conditions to lead to the outbreak of the fire. So maybe we could say: the cause is necessary and sufficient *under the circumstances* for its effects. In limiting cases, Mackie admits, a cause may even be necessary and sufficient on its own, though: causes are, as he puts it, *at least*, **INUS** conditions for their effects (1993, 36).

C4.P16 Mackie's proposed analysis has received a wide range of criticisms. These criticisms were in part inherited as

criticisms of the original Humean analysis that Mackie was trying to refine, and in part specifically targeted at Mackie's own suggestions. Let us briefly consider here three difficulties that were raised against Mackie's account.

C4.P17 The first concerns the distinction between causes and background conditions. An obvious consequence of Mackie's analysis is that every event has many causes. (As we noted, this makes the analysis in one respect attractive.) For just as the short circuit is sufficient and necessary for its effects given the other circumstances (including the presence of oxygen), so the presence of oxygen is sufficient and 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, however, resist the idea that things other than events (like the short circuit) are causes, and say that other factors can only be "background" or "enabling" conditions. If they are right, Mackie's proposal, which rules out this distinction, has an obvious problem. However, in Mackie's defense, it must be pointed out that it is very hard to apply the "cause-versus-background condition" distinction to particular cases without stretching the facts to fit the theory, or making ad hoc stipulations. When we talk about, say, "*the* cause of Caesar's death," this ought not to be taken literally, at face value. Caesar died from a massive loss of blood, but he also died because he was stabbed, and because his bodyguard was not around to protect him that day, and because he was ambitious and wanted to become king of Rome. Is there any reason to think that only one of these things is "the real cause" of Caesar's death, or that only a few of them are "the real causes"? That the absence of his bodyguard was not a proper cause, but only a background or "enabling" condition that facilitated

that Caesar would be killed by the assassins, but did not itself cause his death? Furthermore, one may be skeptical that an in-principle distinction between causes and background conditions has to be part and parcel of any viable theory of causation, for what we consider as background conditions—rather than causes—in a particular case depends, primarily, on our interests or expectations concerning the "normal state of affairs" in situations like the one at issue, and these expectations and interests change all the time. This suggests that the distinction between causes and background conditions is a pragmatic one, rather than a metaphysical one. Therefore, even if Mackie's theory has no good way to distinguish between background conditions and causes, this need not be a decisive objection against it.

C4.P18 A second difficulty raised against Mackie's account concerns the relation between causation and determinism. Mackie's account of causes as **INUS** conditions presupposes that there is a sufficient condition for the occurrence of the effect (even though this condition will encompass more than just individual causes taken singly). Positing sufficient conditions is unproblematic in cases of deterministic causation (i.e., where the effect *must* follow the cause); however, in nature there are also cases of indeterministic causation, or so many believe, such as quantum mechanical processes (where the probability of the occurrence of the effect is significantly raised by the presence of the cause, but the effect is not necessitated). When an effect has only indeterministic causes, there will be no **INUS** conditions for this effect, simply because there is no set of previous conditions which absolutely ensured that this effect would take place. (Even a combination of causal factors, in such cases, merely *raises the probability* of the occurrence of the effect.) Mackie's

conception of "cause," however, rules out indeterministic causes from the start.

C4.P19 Finally, a third difficulty raised against Mackie's account concerns the direction of causation. It is usually thought that causation has a direction, "from" the cause "to" the effect. This directedness expresses itself in a formal feature of the causal relation, namely that it is nonsymmetric: if *x* causes *y*, then (in standard cases at least) *y* does not cause *x*; or at the very least, it does not follow that *y* causes *x*. One well-known challenge for those who attempt to define "cause" in terms of sufficient and necessary conditions is how to account for this directionality. Accounting for the directionality of causation is evidently a problem if one assumes that cause just is a necessary *and* sufficient condition for the effect—since if *x* is necessary and sufficient for *y*, then it follows that *y* is necessary and sufficient for *x*, too. But the general problem is inherited by accounts like Mackie's, because, as we saw, Mackie requires that if "A causes B" then A is at least an **INUS** condition for B, and allows that A can be necessary and sufficient for B on its own. The standard Humean way to account for the direction of causation is to use the direction of time (recall Hume's formulation: "followed by another"). However, this strategy will not work when we look at cases of simultaneous causation, since here cause and effect happen at the same time. (And Mackie cannot appeal to the direction of time to underpin the directionality of causation, because he accepts the in-principle possibility of backward causation, where an event later in time causes an earlier event [1993, 51].)

C4.P20 David Lewis's theory of causation[5] promises to resolve at least the second and the third difficulty raised against

C4.N5 5. For Lewis's original views see the essays collected in his 1986b; for a

Mackie's proposal. Lewis follows quite closely Hume's second definition of "cause": "where the first object had not been, the second never had existed," and provides a counterfactual analysis of (certain paradigm forms of) causation. As we saw in section 3.4, Lewis famously used the "toolbox" of possible worlds and the notion of comparative overall similarity of possible worlds to account for the truth conditions of counterfactual conditionals.

AU: Please do confirm the set character is fine as in the proof in this paragraph.

When "A $\Box\!\!\rightarrow$ C" is true, when can say that "C counterfactually depends on A."[6] Using this notion of counterfactual dependence, which is a relation between propositions, Lewis defines *causal* dependence, which is a relation between particular events. Where "c" and "e" are terms denoting particular events (e.g., "the assassination of the Archduke Ferdinand," "the First World War") and "O" is a predicate of events, meaning "occurs," and "~" is negation, causal dependence can be defined as follows:

e causally depends on *c* iff:

(1) Oc $\Box\!\!\rightarrow$ Oe; and

(2) ~Oc $\Box\!\!\rightarrow$ ~Oe

If *c* and *e* are actual events, then (1) is true because of the stipulation that the actual world is always the closest world to itself. Since *c* and *e* actually exist, there is a c-and-e world that 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. Clause

AQ: Should "c" and "e" be italicized in the highlighted phrases as well?

critical reappraisal see his 2004.

6. For the version of Lewis's account and the terminology we are following here, see Lewis (1993).

(2) is therefore the one which is the real heart of Lewis's counterfactual analysis of causation: 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*.

C4.P26 Causal dependence is not yet the same as causation, though: Lewis says that causal dependence between actual events implies causation, but not vice versa. This is because 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* . . . and so on, where *d* depends on *c, e* depends on *d*, et cetera. Then *c* is a *cause* of *e* when there is a causal chain leading 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*. Imagine, for instance, that your mother woke you up this morning, so that you would get the bus and get to your philosophy class on time. If you had not been woken up by your mother at 8 a.m., you would not have been at the bus stop by 8:25 a.m., but only at 8:40 a.m., and would therefore not have caught the bus at 8:25 a.m.; and if you had arrived at the bus stop only at 8:40 a.m., there wouldn't have been another bus that would have got you to class on time. So, your mother waking you up, by causing you to arrive at the bus stop in time, caused you to get to the class on time. But you could have come to the class on time even if your mother had not woken you up early; if, for just that eventuality, your father had already promised to give you a lift to school in his car. (His promise, however, was restricted to the case in which your mother would not have woken you up in time: once she did wake you up and you were on your way, he was no longer available, for example, to pick you up from the

bus-stop and drive you to the class.)[7] So getting to the class on time did not causally depend on your mother's waking you up early enough. Lewis thus defines causation as the *ancestral* of the relation of causal dependence. The ancestral of a relation R is that relation that 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* if *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. While "*x* is a parent of *y*" is not transitive (your grandmother is a parent of your parent, but not your parent herself), "*x* is an ancestor of *y*" is. The relations "*x* causally depends on *y*" and "*x* is a cause of *y*" have the same structure.

C4.P27 How does Lewis's proposal fare with regard to the second and third problems for Mackie's analysis? With regard to the second problem, Lewis's account is not committed to the idea that all causation must be deterministic. As long as we can make sense of comparative similarity in cases where only probabilistic connections between events are at stake, Lewis's analysis of causation will work for such cases, too. With regard to the third problem, to account for direction of causation in a way that does not just posit that the direction of causation *is* the direction of time, Lewis has tried to show from considerations about comparative similarity among possible worlds that the direction of time does not have to be put in "out of the blue," but falls naturally out of the analysis of counterfactual dependence. (The key idea, which we cannot go into in detail here, is that in our world, events have

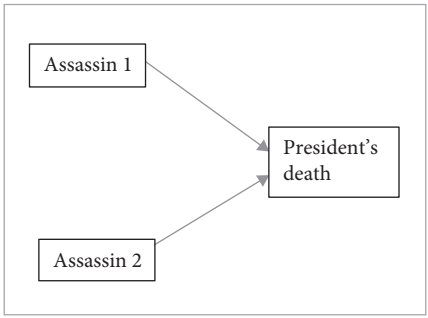
C4.N7 7. This bit of the example is important because it ensures that your being at the bus stop at 8:25 a.m. was, under the circumstances, necessary for your getting to school.

more effects than causes, and that, as a result, worlds with the same or very similar pasts are bound to be more similar to one another than worlds with the same futures.) Not everyone has been persuaded by this answer to the third difficulty raised against Mackie's account, though, and it is also to be noted that Lewis's response does not work for cases of simultaneous causation (where the direction of causation obviously cannot be the direction of time).

C4.P28 However, Lewis's proposal faces some additional problems. To begin with, there seem to be many cases of counterfactual dependence between events that are not cases of causation or causal dependence (see, e.g., Kim 1993). Assume, for instance, that Jim has promised Jane not to drink alcohol tonight, and this was the only promise Jim has ever made to Jane. If Jim did drink alcohol tonight, he thereby broke the promise. Also, given that this was his only promise to Jane, if he hadn't drunk alcohol tonight, he wouldn't have broken his promise to her. So both conditions for causal dependence, in Lewis's analysis, are satisfied. But Jim's drinking alcohol didn't *cause* him to break the promise to Jane: his drinking *constituted* a breaking of the promise. Generalizing this point, philosophers like Jaegwon Kim have objected to Lewis's account that causation is only one of many kinds of dependence relation that can underlie the counterfactual conditionals which Lewis takes to be the core of causal dependence. So how can we pick out the causal dependence relations from the others? (We cannot pick it out, obviously, by saying that it is a dependence that rests on causal connections, if we want to provide an analysis of causation that is not circular!)

C4.P29 In addition, there are some types of case involving causation that Lewis's analysis does not seem to cover, in particular,

cases of causal overdetermination and pre-emption. In overdetermination cases, you have two sufficient causes, such that each of them could be absent (singly), and the effect would still obtain. Assume, for instance, that two assassins, independently from each other, shoot President Kennedy, such that their bullets enter his heart at exactly the same time. Arguably in such a case, if the first assassin had not fired, the second would still have killed President Kennedy. And the same applies, mutatis mutandis, to the second assassin. So none of their shots was necessary, and for each the counterfactual "If the shot had not been fired, the president would not have died" is false. Nonetheless, we would usually consider both shots as causes of the president's death.[8]



C4.P30 Pre-emption cases are trickier. In pre-emption cases, event A causes event B, but if A had failed to obtain or to have a causal influence, there was a "back-up" cause waiting "in the wings" that would have caused B if A had failed to obtain or to have a causal influence. In such cases, even if A is a cause of B, B would still have obtained if A had been absent.

C4.N8 8. Lewis himself does not want to address these cases because, as he puts it, he lacks "firm naïve opinions about them" (1986b, 171 fn. 12). But many philosophers did have stronger positive intuitions about these cases.

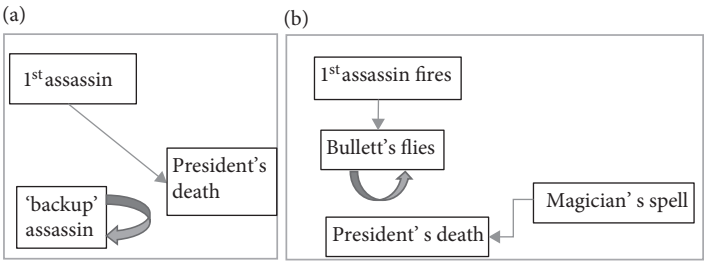
Imagine that one of the two assassins shoots the president, hits his mark and kills him instantaneously. The second assassin was just there as a "back-up," who fired his shot two seconds after the first, to "make sure" in case the first assassin's bullet was to go wrong (or in case the first assassin had changed his mind and not fired at all). Let us assume that the second assassin was a dead-certain shot, and his bullet would have certainly have killed the president; however, when it arrived, the president was already dead (since already the first bullet had been lethal). In this case, clearly it was the first assassin's shot that killed the president (since the second assassin's bullet arrived too late). Nonetheless, it is not true that if the first assassin hadn't shot, the president wouldn't have died (since, in that case, the second assassin's bullet would have killed him). Thus the analysis says—incorrectly—that the first assassin's shot did *not* cause the president's death.

C4.P31

Lewis's own response to pre-emption cases appeals to his definition of causation in terms of a chain of causally dependent events. The first assassin's shot causes the president's death because there is a chain of events between this shooting and the death (the bullet flying out of the gun and through the air, hitting the president's head, etc.). 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 (remember the going-to-class case discussed earlier). So, the fact that there is a back-up does not prevent the shot from causing the death. However, this response strategy only works when the pre-emption case involves a chain of intermediate causal steps, and causal dependence only fails for the earlier steps in the chain. But could the possible pre-empter not be waiting in the wings just at the very last (or the only) step in the chain? Imagine, for instance,

AQ: Please note that as per OUP standards terms like "above" "below" etc., are avoided. Is "discussed earlier" fine here?

that a magician can directly cause the president's death by casting a spell. If the magician casts the spell immediately before the bullet hits the president and his saying the spell directly kills the president, the only cause of the president's death is the spell, not the bullet (which arrived too late to kill the president). Nonetheless, that the bullet was already underway made it necessary that the president would die anyway. This seems to be a case of pre-emption not amenable to Lewis's response strategy, since the causal chain from spell to death involves no further intermediate steps and causal dependence does not hold with regard to the only step in the chain which led to the death.



C4.P32 Even refined neo-Humean theories such as Mackie's and Lewis's thus face serious difficulties (of which we have only been able to discuss a very small selection here). Such difficulties have motivated a shift of interest among metaphysicians: an increasing number of philosophers have come to think that these difficulties make it worthwhile to look back at the Aristotelian conception of causation that Hume had tried to demolish, and are exploring whether it might be fruitful to start afresh from there. That's what we will do in the following two sections of this chapter.

C4.S3

4.3 ARISTOTLE'S THEORY OF CAUSATION

C4.P33

Aristotle famously distinguished four types of cause.[9] Take a particular statue made by Fidias, the statue of Athena on the Acropolis in Athens. Aristotle's view is that the marble from which it is made is its *material* cause; the form (of Athena) is its *formal* cause; Fidias who sculpted it, its *efficient* cause; and the purpose for which it was made, for example, to adorn the Acropolis, its *final* cause. Today we find it somewhat odd to talk of each of these four things as "causes," since we have become used to reserving the term "cause" to the third kind of Aristotelian causes, namely efficient causes (and, at best, regard the other things as "explanatory factors"). But this apparent oddness is mostly due to the fact that our conception of causation radically changed in the seventeenth century, as part of the rise of the natural sciences and their paradigms of explanation. For Aristotle himself, by contrast, as well as for other philosophers of antiquity and the middle ages, there are more types of cause than the efficient cause,[10] and he has no problem seeing all his four causes as genuine "causes," for which there should be a unified metaphysical account. The account he offers is, simply put, that all causes are powers, and causation, of all kinds, is the exercise or manifestation of mutually dependent causal powers. The mutual dependence among powers is grounded in the fact that for Aristotle they are (monadic) properties of a special type that he calls "relatives" (as we saw in chapter 2; remember the passage we

C4.N9 9. This section draws on Marmodoro (2007).

C4.N10 10. See, e.g., Sorabji (1980, 26–44); Frede (1987, 125–50).

quoted from *Categorie*s 7b6–7: "if there is no master there is no slave either"). The relevant powers are mutually dependent not for their existence but for their exercise: any agent requires a patient upon which to act.

C4.P34 To get a better idea of this theory, let us focus on the analysis Aristotle gives of the causal interaction (of the efficient causation type) between mover and movable in his discussion of *kinesis* (change, motion) in *Physics* III. Aristotle's definition of motion (see, e.g., 201a9–10; 201a27–9; 201b4–5; 202a13–4) is fairly broad[11] and includes what we would consider uncontroversial and paradigm instances of causation, such as building, heating, and doctoring. Thus, Aristotle's account of the mover-movable interaction will serve as a good model for explicating more generally his metaphysics of causation. What are the key elements of a causal interaction such as the one existing between a mover and a moved? Aristotle begins with a programmatic statement: accounting for motion does not require appealing to any new, primitive category of being (*Physics* 200b32–201a3); instead, Aristotle will use some of the key ideas he has introduced elsewhere without adding anything more to his ontological "stock" in order to explain motion and change. These key ideas are: forms, the privation of form, the substratum of change, and the distinction between being in potentiality and being in actuality. He describes efficient causation as the "transmission" of a form from the mover to the movable:[12]

C4.N11 11. Aristotle's definition allows for a great variety of cases to come under the mover–movable relation, including such cases as aging, which we would consider as untypical cases of causation at best.

C4.N12 12. This is, however, as we will see in a moment, only a figurative description; there is no item that gets literally passed on.

C4.P35 > The mover will always transmit a form, either a "this" or such or so much, which, when it moves, will be the principle and cause of the motion, e.g. the actual man begets man from what is potentially man. (*Physics* 202a9–12)

C4.P36 So in general terms, in causal interactions between substances the agent transmits and the patient receives the form; the same form is involved "on both sides," only that they relate to this form differently. The agent's transmitted form is the cause; the privation of the form in the patient is what allows for its reception, and the physical process facilitating the transmission of the form is the substratum of the causal change (e.g., in building, the movements of the hands of the builder facilitate the transmission of the form of the house to the construction materials; for heating, contact facilitates the transmission of the form to the object heated). Describing (figuratively) causation in terms of "transmission of the form" plays an important role in Aristotle's account of causation. In particular, it enables him to address the question of whether causation follows the order of time, with causes preceding their effects. If not, is it mere convention that in the causal interaction between the teacher and the pupil, teaching is the cause and learning the effect? Or is there an underlying metaphysical principle for the determination of the cause and the effect? Aristotle writes,

C4.P37 > A man may have hearing and yet not be hearing, and that which has sound is not always sounding. But when that which can hear is actively hearing and that which can sound is sounding, then the actually hearing and the actual sound come about *at the same time* (these one might call respectively hearkening and sounding). (*De Anima* 425b26–426a1, emphasis added)

C4.P38 Aristotle is thus clear that actual causes do not precede their actual effect in time; teaching and learning (by being taught) cover the same time span. The passage just quoted also shows that Aristotle conceives of causation as the exercise of causal powers that get mutually activated: the power of sounding in the example, and of hearing. The power to teach is in the teacher before she engages in actual teaching (and even if she may never engage in actual teaching), and so is the corresponding passive power in the learner, but the actualization of these potentials is one and the same, hence there is complete overlap in time between them.[13] In the *Physics* Aristotle gives a metaphysical account of his key claim that the actualization of the active and the passive powers is *one and the same*, which is what also explains their temporal coincidence. He writes:

C4.P39 > Motion is the fulfillment of the potentiality of the movable by the action of that which has the power of causing motion . . . A thing is capable of causing motion because it can do this, it is a mover because it actually does it. But it is on the movable that is capable of acting. *Hence there is a single actuality of both of them alike.* (*Physics* 202a13–18, emphasis added)

C4.P40 Aristotle's claim will look, at first sight, puzzling, but he explicates it when he examines some alternative accounts of what happens to the mover and the movable in causation in a dialectical puzzle in *Physics* III 3, 202a21–b5.[14] In the course of this discussion, Aristotle rejects some of the possible

C4.N13 13. It is important to note, though, that this complete overlap does not raise problems for accounting for the direction of causation. For the direction of the transmission of the form is asymmetric, and that determines which is the agent and which the patient in the causal interaction.

C4.N14 14. See Marmodoro (2007, 207, 230–31).

alternatives to the "single actuality" claim, while developing the rationale behind his own explanation of the "oneness" of agency and patiency. In brief, Aristotle considers two possibilities: that the two actualities, of the mover and the movable, are different, or that they are one and the same. The first option seems to lead to a dilemma. If the two actualities are different, either they both occur in one of the two subjects (namely, either in the mover or the moved), or one occurs in each. If both actualities occur in one subject, then whichever has both actualities in it will change in two different ways in relation to one form. For example, when we take the case of A teaching B, then one person would come to be both teaching and learning at the same time by teaching, and this is, to Aristotle, seems absurd (and, presumably, we can follow him in this assessment when we exclude those cases where someone teaches herself). If, on the other hand, the actuality of the mover is in the mover, and the actuality of the movable is in the movable, then either the causal agency of the mover will impact only on the mover itself, and not on the movable (so the teacher will teach herself, but not the pupil), or it will impact on nothing (the teacher will teach nobody), in which case the mover is not a mover in actuality. Given that the assumption that the actualities are different leads to untenable results, it appears that the actualities cannot be different and must be one and the same. But this latter result is no less absurd, because agency and patiency cannot be the same. Aristotle's own solution is that the two actualizations, of the agent's and of the patient's respective powers, are different, but also one, because interdependent, and they both take place in the patient. Immediately after developing the dilemma we just looked at, Aristotle states his own position on the issue:

C4.P41

> Nor is it necessary that the teacher should learn, even if to act and be acted on are one and the same, provided that they are not the same in respect of the account [*logos*] which states their essence [*to ti ên einai*] (as raiment and dress), but are the same in the sense in which the road from Thebes to Athens and the road from Athens to Thebes are the same, as has been explained above. (*Physics* 202b10–14)

C4.P42

Thus, for Aristotle, even if both actualities are, in one sense, the same, they are different "in respect of their *logos*." There has been much discussion in the literature concerning how to understand this last crucial term, *logos*, but the context makes it clear that it must refer to the definition of the respective natures. For Aristotle uses in the passage the technical expression he has coined for essence (*to ti ên einai*, which we already know from the discussion of Aristotle's essentialism in section 3.2.). You might think that this reading makes Aristotle's position even more puzzling. For it is a cornerstone of Aristotle's essentialism that if the essences of two things are of different kinds (say a wolf and a rabbit), the two things in question must be essentially and numerically different. What it is to be an agent is different from what it is to be a patient; their definitions are different (202a20, 202b22), and with them, their kind (202b1). Nonetheless, although the definitions stating their essences (202b12) are different, Aristotle maintains, "to act and to be acted on are one and the same" (202b11). The example he offers helps us to better understand what he means by this. The route from Athens to Thebes and the route from Thebes to Athens are, in one sense, "one and the same": for these routes are embodied in the same stretch of the road. In the same way, the ground of the respective actualizations of agency and patiency is one

and the same, notwithstanding that agency and patiency are actualizations of essentially different powers. Aristotle states this explicitly when he writes:

C4.P43

> To generalize, teaching is not the same in the primary sense with learning, nor is agency with patiency, but that to which those belong [*scilicet*] is the same for both, namely the motion; for the actualization of this [teaching] in that [learning] and the actualization of that [learning] through the action of this [teaching] differ in definition. (202b19–22)

C4.P44

The two powers or potentialities in question differ in type, thus their actualities differ in type, too. But the actualities of the two potentialities (for teaching and for learning) are fulfilled in the motion that is the common activity that actualizes them both. The two potentialities can occur in actuality only together, and neither can happen without the other. The teacher is not teaching if the learner is not learning, and the learner (i.e., instructee) is not learning (being instructed) if the teacher is not teaching. The oneness of the activity grounding them accounts for the interdependent actualization of the cause and the effect. At the same time, the two essences or forms that the activity bears preserve the bipolarity of the causal interaction. Lastly, as we will see in the next section, the fact that the activity that actualizes both agency and patiency is located in the patient grounds the distinction between agent and patient, and between the active and passive powers which are involved in the causal transaction.

C4.S4

4.4 ARE THERE *ACTIVE* AND *PASSIVE* POWERS INVOLVED IN CAUSATION?

C4.P45

The Aristotelian view we discussed in the previous section has many attractive features. And, as we have noted at the end of section 4.2, the problems which neo-Humean accounts of causation still face have motivated a growing number of philosophers to look back to Aristotle's model of causation and to take over at least some crucial elements of it. Contemporary neo-Aristotelians do believe that one can provide an account of causation in terms of the actualization of potentialities (as Aristotle would have put it) or the exercise of powers (as we would put it). This Aristotelian idea, they hold, allows us to revive the idea that causation involves "production," and that the effects "derive" from their causes, which had been lost by the Humean tradition (Anscombe 1993, 92). Different approaches to developing this have been tried out; they can be classified in different ways. One important distinction depends on whether causation is accounted for in terms of the coming together or "coactivation" of different powers, *active* and *passive*; or whether causation need involve the manifestation of one kind of power only, namely an active power.[15] Another important distinction can be drawn according to what these approaches take the relata of causation to be: powers themselves, events, or objects? And yet another distinction can be drawn according to whether

C4.N15

15. For the first view see, e.g., Marmodoro (2007) and Mumford and Anjum (2011); for the second, O'Connor (2009) and Mayr (2011).

one believes that all instances of causation, or only some, can be understood in terms of the activation of powers.[16]

C4.P46 Many—but not all—power theories of causation adopt the distinction between active and passive powers in some form or other. What makes this move attractive is that it promises to provide an explanation for the direction of causation, which, as we have seen in section 4.2, raises special difficulties for neo-Humeans. Some philosophers, though, have rejected the distinction between agent and patient in causal transactions, and argued that all causal transactions are symmetrical, in the sense that all involved transaction partners undergo changes equally. Take, for instance, the following passage by John Heil:

C4.P47 > The received view of causation might lead you to think that the water and the salt are related as agent and patient: the water, or maybe the water's enveloping the salt, causes the salt to dissolve. Perhaps the water possesses an "active power" to dissolve salt, and salt, a complementary "passive power" to be dissolved by water. But look more closely at what happens when you stir salt into a glass of water. Certain chemical features of the salt interact with certain chemical features of the water. . . . This interaction is, or appears to be, continuous, not sequential; it is, or appears to be, symmetrical. (Heil 2012, 118)

C4.N16 16. In the literature, this last question is sometimes dealt with under the heading of whether all kind of causation is of one kind, e.g., substance-causation (Lowe 2002), or whether there are irreducibly different kinds of causation, e.g., some kinds where the causation consists in the activation of a causal power of a substance and others which run along more or less Humean lines. Some so-called agent-causalists in the free-will debate go for the latter option, e.g., O'Connor (2000), and Clarke (1996).

C4.P48 Much-discussed cases, that seem to display this symmetry in a particularly striking way, are cases where two cards lean against one another and sustain each other in their position (Heil 2017), and cases where a cube of ice cools the lemonade in a glass while being itself warmed up (and dissolved) (Mumford 2006). Cases such as the two mutually sustaining cards are very common in nature and show, according to Heil, that there is no real metaphysical distinction between agent and patient, and hence between the manifestation of active and passive powers. There is only one single manifestation of both, the fact that we call it "dissolving" or "being dissolved" (in the case of salt in water) is a perspectival matter based on how we look at events and what we are interested in.

AQ: No Mumford 2006 in biblio; pls. add.

C4.P49 Heil's argument is not compelling, though. Let us grant, for the sake of the argument, that there are some cases of symmetrical interactions, like the case of the two cards leaning against one another. Even then it would be wrong to infer that *all* cases work like that. Let us return to our original case of the water dissolving the salt. Even if both the salt and the water change during the process, Heil's claim that the interaction is fully symmetrical is unwarranted. For there are still important differences between what the salt and the water do to each other. For instance, polarized water molecules break the bond between the negative chloride ions and the positive sodium ions, while salt molecules do not break any water molecule. The chemical reaction is asymmetrical. The fact that water dissolves salt but not vice versa is scientifically explicative: if we say that water dissolves salt we are communicating scientific knowledge, and more specifically knowledge about the relevant causal process.

C4.P50 What then differentiates active and passive powers, if their distinction is what underpins the asymmetry and

directionality of causality? Their distinction can either be thought to be an absolute one, such that a power is (generally, by nature) either a passive or an active one; or a relative one, such that powers take on an active or passive "role" in causation depending on how they are involved in causal interactions, as Aristotle for instance thought. For Aristotle, a power is *active* if it causes a change in another power (or its bearer); and is passive if it (or its bearer) undergoes change. (Even if both powers change, one can change more or more radically than the other, as in the example of water and salt: so we can still compare degrees of undergoing or causing change even then.) An alternative way to distinguish between active and passive powers focuses on the (relative) dependence of the manifestation of a power on the external circumstances. A power is comparatively more active, on this view, the wider the variety of circumstances under which it can manifest, and the less it has to be activated under these circumstances. This is one way to capture the idea that causation is tightly connected to the explanatory contribution something makes to an effect, and this explanatory contribution is relatively greater for a power if the relative explanatory contribution of the external circumstances are relatively smaller.[17] In both suggested ways of distinguishing active from passive powers, it is important to note, the attribution of active and passive roles is not merely based on epistemological considerations, but on metaphysical ones.[18]

C4.N17 17. See, e.g., Harré and Madden (1975); Mayr (2011, chap. 8).

C4.N18 18. Clearly, the attribution of active and passive role to powers in a causal interaction is neither easy nor unequivocal. However, this doesn't mean that the distinction, in general, makes no sense, and this does not exclude that in some cases the active and passive powers are clearly distinguishable.

C4.S5

4.5 MENTAL CAUSATION

C4.P51

It seems a trivial truism that mental occurrences can have causal effects; and equally it seems obviously true that mental occurrences can cause not only other mental occurrences, but physical ones as well. Imagine yourself thinking hard about the solution to a mathematical problem ("What is 246 + 874?") and suddenly grasping the solution ("Ah, it's 1120!"). Your thinking "Ah, it's 1120!" is a mental event. It can not only cause you to have other thoughts (e.g., your thinking "So the overall sum is 5698"), but can also have physical effects (e.g., your writing down "1120" on your exam sheet). Or imagine yourself feeling a sharp pang of pain as you touch an extremely hot cup of tea. Wouldn't we naturally assume that it is your pain that causes you to cry out loud (which is a physical occurrence)? Again, when you intend to do something, this intention, it seems, can clearly impact causally on what you are doing: unless you are weak-willed or somehow prevented, your intention will make you act in certain ways.[19] These three cases, and others like it, seem to be clear cases of causation. So it seems obvious that mental occurrences can have both mental and physical effects.

C4.P52

Nonetheless, since the time of Descartes the problem of mental causation has been haunting metaphysicians. In Descartes's own system the problem arises from his mind–body dualism, which includes a strict distinction between

C4.N19

19. It is disputed whether this is best understood as your intention causing you to act in certain ways. But because we do not follow the traditional picture of event-causation here, we will set this problem aside, because the intention can at least manifest itself in your actions, which is enough for its being causally relevant.

purely mental entities (souls), which are characterized by the property of consciousness, and physical entities (bodies), which do not have any mental properties. Descartes's dualism made the question of how these two types of substance, mind and body, could interact particularly pressing for him—while his philosophical assumptions made this question almost unanswerable. To begin with, the mind was, for Descartes, not localized in any part of the body, which left it completely mysterious how it could impact on the latter. Famously, Descartes thought that there was a specific organ in the body, the pineal gland, that allowed the soul to act upon the body, but positing that there was such an organ hardly solved the general problem of how a purely mental entity could interact at all with another object completely different from it. (Several philosophers indeed despaired that providing an answer to this question was possible; most notably Malebranche, who thought that every apparent interaction between soul and body was mediated by divine intervention. This view is known as Occasionalism.)

C4.P53 The problem of mental causation as it arose for Descartes was directly tied to his view of substance dualism: that is, the view that mental and physical substances formed two exclusive kinds of substances that did not share any of their essential properties. Substance dualism has become very much a minority position today, however, and its demise has largely been due to its inability to explain causal interactions between mind and body. Nowadays, most philosophers accept some form of physicalism, according to which—minimally—*all* substances have physical properties and are physical entities. But the problem of mental causation has returned in different forms—and, indeed, it seems safe to assume that as long as mental and physical items, be they substances, properties, or

events, are considered to form exclusive sets—this problem will haunt philosophers in some form or other.[20] Even if you believe that all substances are physical substances, you will still have to assume that some of substances have mental properties and states, and are subject to mental occurrences, too. We human beings, for instance, have mental properties and states: we have beliefs and desires, we sometimes feel joy, or acute pangs stabs of pain. How are these properties, states, and occurrences related to our physical properties and states, and the physical events in which we are involved?

C4.P54 One way to answer this question is to hold that mental properties and physical properties are identical; that is, that for every mental property P there is a physical property P' (which may be a complex, or disjunctive, physical property), such that P is identical with P', and that instantiating P is the same as instantiating P'. There would then be no specific problem of mental causation any more: if mental properties were identical with physical properties, their causal relevance would be explained just in the very same terms as the causal relevance of other physical properties. This view is held by reductive materialists—but it is not a view that many contemporary philosophers still subscribe to. Influential arguments (such as Hilary Putnam's [1975] multiple realization argument that the same mental states can be realized by many different physical set-ups) have convinced

C4.N20 20. What makes an event or property a mental or a physical one? This question is a tricky one and not all participants in the debate give the same answer. Two features that have often been used to characterize mental items are consciousness (Descartes) or intentionality (Brentano). For a rough-and-ready characterization we will use the disjunction of both: X is a mental property/event/state if it essentially involves consciousness or intentionality.

most of them that mental and physical properties cannot be identical. If both properties are not identical, though, then types of mental states cannot be identical to types of physical states, either, and being in mental state P cannot be exactly the same thing as being in physical state P'. For plausibly, being in a certain state is (at least in part) defined as instantiating a certain property or standing in a certain relation, such that two states cannot be identical to one another unless the properties involved are the same, too. Given that most of these philosophers still subscribe to a version of physicalism, their views are usually called versions of "non-reductive physicalism."[21]

C4.P55 "Non-reductive physicalism" is, arguably, the mainstream position in philosophy of mind today. But if mental and physical properties/events/states are not regarded as identical, the problem of mental causation rears its head again. One of the most influential contemporary versions of the problem has been formulated by Jaegwon Kim, with his so-called exclusion argument (1998, 30), which targets specifically non-reductive physicalists. Physicalists, says Kim, must assume that there is some connection between physical and mental properties, and one widely accepted way to think about this connection is in terms of supervenience. While there are differently notions of supervenience, the core idea is that properties of type A supervene on properties of type B if there can be no difference in the A-type properties of

C4.N21 21. Of course there are other options, too, in addition to believing that mental and physical properties are identical or believing they are fully separate. You might, for instance, think that one and the same property has both mental and physical aspects. But for ease of presentation, we will set these further options aside here.

an object, state, or event, without some difference in the B-type properties (either of the object, state, or event itself or its "surroundings," or history). Such a supervenience assumption is very common, for example, with regard to the relation between moral and nonmoral properties: two actions cannot differ in their moral evaluation, it is widely assumed, unless there is some nonmoral difference between them as well. The same, Kim argues, can also be supposed to hold with regard to mental and physical properties:

C4.P56 > Mental properties *supervene* on physical properties, in that necessarily any two things (in the same or different possible worlds) indiscernible in all physical properties are indiscernible in mental respects. (Kim 1998, 10)

C4.P57 The instantiation of any mental properties must thus have an underlying "basis" in the instantiation of some physical properties, which is usually called the "supervenience base." By itself, the supervenience thesis does not imply any claim that the physical properties are more basic than the mental ones. But most of the philosophers Kim's argument was originally intended to address, being physicalists, would also accept that the physical properties are more fundamental, and that it is due to their instantiation that the mental properties are instantiated as they are (and not vice versa).

C4.P58 A second key assumption of the "exclusion argument" is the thesis of the "causal closure of the physical realm." While there are different formulations of this thesis, the core idea is that the causal laws that govern the physical realm only allow for causation of physical events by other physical events, since for every physical event there is we can already provide a full causal explanation of it in terms of *physical* causes. Kim

holds that this assumption of "causal closure" is another key tenet of physicalism that one cannot give up as long as one subscribes to at least a weak form of physicalism:

C4.P59

> If you reject this principle, you are ipso facto rejecting the . . . completability of physics—that is, the possibility of a complete and comprehensive physical doctrine of all physical phenomena. . . . It is safe to assume that no serious physicalist could accept such a prospect. (Kim 1998, 40)

C4.P60

Now, it is not difficult to see that these background assumptions (that physical properties are fundamental and of causal closure of the physical) generate a fundamental difficulty for the possibility of mental causation. Let us begin with the question of whether, on this picture, mental events can cause physical effects. The answer is clearly no. If one reads the causal closure principle as generally excluding nonphysical causes for physical effects, it directly rules out this possibility. But even if one gives the principle a weaker formulation (that is, allowing for the existence, in principle, of nonphysical causes in addition to physical ones), causation of physical effects by mental causes will be highly problematic: for, by causal closure, the physical effects will already have physical causes, too, so what work is there left for the mental causes to do? To make matters worse, when the mental event appears as a putative cause of a physical event, there will be an obvious contender for the former's causal role, namely the (physical) supervenience base of this mental event. All causal influence that the mental event could have will necessarily reduce to the influence of this physical base—or so it seems. The case of mental-to-mental causation seems, at first, more promising, since the closure principle doesn't tell

us anything about the causes of mental events. However, as Kim argued, even here there is fundamental problem. When a mental event A causes another mental event B, B will also—according to the supervenience thesis principle—have a physical supervenience base B*. And, when B supervenes on B*, Kim claims, it is plausible to assume that one can only cause B by causing its supervenience base to obtain, that is, by causing B* to obtain. But since the obtaining of B* is a physical event, A's causing B* would be a case of mental-to-physical causation—and, as the first part of the argument was meant to show, this kind of causation is ruled out by the causal closure principle. If we take this principle seriously, we will not take A to be the genuine cause of B* (and, by extension, of B itself), but, instead, A's supervenience base A*. Could one, however, not say that both A and A* cause B, because A* causes B via causing A? No, says Kim, because the supervenience relation is not a causal relation. Thus we do not have a causal chain going from A* *via* A to B* and B, but we have only a direct causal link from A* to B*. Attributing any causal role to A (e.g., going directly to B) would be a merely gratuitous extra—and would make the case one of causal overdetermination, where B would have two independent causal histories which would both fully explain its occurrence. And it is implausible to think mental events and their physical supervenience bases would play such independent causal roles, which would lead to causal overdetermination.

C4.P61 Kim's argument has encountered strong criticism—unsurprisingly, since, if successful, it would destroy too much of our commonsense picture of the world. Indeed, it would not only undermine mental causation (which would be bad enough); it would also (see, e.g., Humphreys 1997) "threaten"

the plausibility of practically all ordinary causal explanations, which are phrased in terms of macro-physical phenomena that are taken to supervene on microphysical ones. Surely, we do not think that ordinary causal explanations such as "the rise in temperature caused the plants to whither" are all false, even though we do think there are underlying microphysical phenomena (of which we yet lack full understanding) on which the rise in temperature and the withering of the plants supervene. Having to reject all such ordinary explanations seems in itself a conclusion that is too hard to accept, and that which gives us strong reasons for rejecting at least one of the premises Kim uses in his argument (e.g., Baker 1993). But which part of the setup of Kim's argument should we deny? One might, of course, deny non-reductive physicalism in the first place. But given the latter's popularity, the premise that has usually been taken to be most problematic is the causal closure assumption. Why should we think that physical effects must have only physical causes, or that they must at least have a full causal history given in terms of physical causes only? We may have independent reasons for accepting some form of supervenience, but causal closure is not an intuitively plausible principle, since it clashes with many of our ordinary causal explanations, where we do explain physical effects by appeal to nonphysical causes. Indeed, seeing how this premise eventually leads us to deny (as long as we accept supervenience) that any phenomena other than microphysical phenomena can be causally effective, it is even doubtful that the robust "realism about science" that is meant to motivate the principle (and which Kim clearly endorses; see preface of his Kim [1998]) really supports it. For such a realism about science should lead us, if anything, to be realist about the subject matters of *all* the

sciences, *not just microphysics*. At least the other natural sciences, such as organic chemistry and biology, should, even for naturalist philosophers, be above suspicion. Of course, you might still worry about how the processes and events specified in these different sciences causally interact, so there is still *some* question to be answered here. But Kim's argument is both specifically based on a "realism about science," and derives from this that mental causation is *impossible* (not merely that it raises additional worries). And if we are genuine realists about science, and not simply realists about microphysics only, this argument will not work, for there will be no reason to believe that in principle, only explanations of events in terms of causally effective microphysical processes can be genuine causal explanations. A more liberal naturalism that takes all the sciences seriously, may thus be a more "mature" successor to the physicalism of the kind Kim subscribes to, which is too deeply rooted in the reductionist programs of the first half and mid-twentieth century, which tried to find a common basis for all the sciences in physics. If such a more liberal naturalism is incompatible with causal closure and supervenience, and causal closure is the more dubious of the two premises, then all the worse for causal closure!

C4.P62 There is also another consideration that speaks strongly against the causal closure principle that Kim proposes, namely that when we focus on event causation, causes should be "proportional" to their effects (e.g., Yablo 1992).[22] The idea behind this "proportionality" response to Kim is the following: when we ask "What caused *x*?," we do not always

C4.N22 22. Yablo (1992, 227) argues that usually, among the various candidates for the cause of an effect, the most proportional candidate should be preferred.

want to know the most specific factor that can be picked out; rather, we want to know the factor that "made the difference," and this can be a relatively coarse-grained one. Imagine for instance that you are wearing a crimson red t-shirt and a bull, on seeing the t-shirt, gets angry and charges you. What was responsible for the anger of the bull (never mind Kim's argument about the causal impact of mental events)? In some sense, obviously, your crimson t-shirt having the color it had. However, your t-shirt is not just crimson; it has a particular shade of crimson, for example, dark crimson. Its being crimson supervenes on its having this particular shade (it could not have had another color without some change with respect to this shade) and, in a way, its being crimson is due to its having that shade. Nonetheless, in response to the question of what caused the bull to be angry, it would be wrong to single out the t-shirt's specific shade as the causal factor. For the bull would have gotten angry *regardless of which particular shade of crimson your t-shirt had*: its being crimson was all it took for the bull to get angry. So, it was the t-shirt's being crimson that made the difference to the bull's getting angry, not its having the particular shade of crimson it had.

C4.P63 This shows that sometimes it is the supervenient property that makes the difference and is rightly singled out as the cause, rather than its supervenience base. Even if every instantiation of a mental property is supposed to have a physical supervenience base, it might therefore still be the instantiation of the mental property that makes the difference, rather than the instantiation of the physical property that underlies it. Assume that one supervenient mental property instantiation can have different physical supervenience bases: then which of these different bases occurred in the particular case under consideration might be completely

irrelevant to the occurrence of an effect. The only relevant factor is that the mental property was instantiated, regardless of what its underlying base was on that particular occasion.

C4.P64 There are therefore several possible ways to respond to Kim's exclusion argument. There are even more options if we adopt a power-based view of causation of the sort we have discussed in sections 4.3. and 4.4. As we have seen in chapter 2, powers are at least partly individuated by what they are powers to do (i.e., their characteristic manifestations). At the same time, that a power's manifestation includes a physical effect does not automatically mean that the power is a purely physical, nonmental one. For the conditions for its exercise, or its exercise itself, may essentially involve mental elements too. Assume you have an active power to bring about a physical effect in the world—for example to move your arm. Need this power be a physical one? Not necessarily, especially if we take the rough-and-ready criterion of "mental" that X is a mental property if it essentially involves consciousness or intentionality.[23] For your power might be a power to move your arm when you want to: then its exercise would essentially involve intentionality and the power would thus count as a mental power. Nonetheless, the effect of the manifestation of this power is something physical: that your arm moves. Already in virtue of this rather trivial feature—that they have characteristic manifestations, while, at the same time, having a physical manifestation does not necessitate that a power is physical one—powers can much more easily bridge the putative divide between mental and physical, thereby undercutting the very setup from which

C4.N23 23. See note 20.

the problem of mental causation arose.[24] Adopting a power-based account of causality thus brings with it the hope that mental causation might be no more problematic than other forms of causation. And given how commonplace a feature of our world mental causation is, this seems just another attractive feature of such account.

AQ: FN 16: Mayr 2016 entry is not in the Reference list. Please add.

C4.S6

4.6 CONCLUSIONS

C4.P65

Causality plays a crucial role in every attempt to explain the fabric of the world. However, as we have seen in this chapter, there is no agreement among philosophers about how to account for causation. We have examined the main alternatives in the debate, tracing them back to Aristotle and Hume. We have seen that an account of causation based on realism about powers has the resources to overcome key issues that afflict Humeanism. The advantages of the former emerged also in relation to the debate on mental causation, where realism about powers would allow us to avoid the influential causal exclusion argument.

*

C4.N24

24. For different ways in which a model of causation based on the realization of powers could turn the tables on Kim's argument see Mayr (2011, ch. 9, and 2016).