

CHAPTER I

Theories of the Mind and Theories of Causation

1.1 Introduction

Explaining how various theories about the nature of mind can accommodate mental causation requires some groundwork. It requires formulating these theories about the nature of mind more precisely. It also requires getting clearer about the nature of causation, which in turn has two aspects: the nature of the relata of causation and the nature of the relation itself. These are the tasks of this chapter.

The aims of the chapter are modest in several ways. I will not attempt to give a complete taxonomy of views about the nature of mind. Instead, I will confine myself to views that are both common and that stand a *prima facie* chance of accommodating mental causation. I will not say much about the comparative advantages and disadvantages of these positions, not least because I do not wish to commit myself to any of them in this book. I will not try to defend a full-blown theory of causation that states necessary as well as sufficient conditions for causation. For one thing, giving such a full-blown theory of causation may well be impossible. For another, it is not necessary for our purposes. Giving a sufficient condition for causation is enough, at least if this condition can be applied in sufficiently many cases of (putative) mental causation. Fortunately, sufficient conditions for causation are easier to find than necessary ones.

The plan for the chapter is as follows. Section 1.2 formulates different theories about the nature of mind. In particular, it defines reductive physicalism, non-reductive physicalism, and dualism. It also defines a version of dualism, naturalistic dualism. Sections 1.3–1.5 discuss the nature of the causal relata and the causal relation. In the context of mental causation, the causal relata are best conceived of as particular or token events, and particular events are best conceived of as being constituted by, and having their identity determined by, triples of an object, a property, and a time (see Section 1.3).

As for the causal relation, we can give a sufficient condition for causation in terms of difference-making or counterfactual dependence: one event causes another if (roughly) the second event would not have occurred had the first event not occurred. The counterfactual conditionals that are used to formulate claims about counterfactual dependence exhibit some logical peculiarities (see Section 1.4). The sufficient condition for causation in terms of counterfactual dependence is subject to a few *prima facie* problems, but they can be overcome by making certain assumptions about how the relevant counterfactual conditionals should be evaluated and by restricting the sufficient condition to suitable kinds of causal relata (see Section 1.5). The sufficient condition for causation in terms of counterfactuals conflicts with the view that causation requires the transfer of a physical quantity from cause to effect. Section 1.6 argues that this conflict should be resolved in favour of the counterfactual condition. It also discusses the requirement that causation should involve an intrinsic connection between cause and effect.

1.2 Varieties of Physicalism and Dualism

Generally, physicalists with respect to X hold that X is physical.¹ Cartesians deny that mental substances are physical and thus fail to be physicalists about mental substances. This view has few adherents today;² I shall follow the mainstream view and assume that all substances are physical. Instead of focusing on substances, I shall focus on mental and physical properties. Physicalists about mental properties hold that mental properties are physical. This claim can be formulated in different ways. The most straightforward reading is that mental properties are identical to physical properties. This is the position called reductive physicalism, defined as follows:

Reductive physicalism: Each mental property is identical to a physical property.

According to reductive physicalism, the property of having a headache is identical to, say, the property of having firing c-fibres:³ to have a headache is one and the same thing as to have firing c-fibres.⁴

¹ Characterizing the physical raises some problems of its own, which I will ignore here. See Crane and Mellor 1990 and Crook and Gillett 2001 for discussion.

² Hart (1988) is an exception. Lowe (1996) endorses a more attenuated substance dualism.

³ Associating pains with c-fibre-firings is empirically questionable, but has a longstanding philosophical tradition, which I follow here. Indeed, I take the liberty of associating c-fibre firings not with pains *per se*, but with headaches.

⁴ As was mentioned in Section 0.2, reductive physicalism was originally called the identity theory and was pioneered by Place (1956) and Feigl (1958). For a recent defence, see Polger 2004.

Since identity entails mutual necessitation, reductive physicalism makes it impossible to have a headache without having firing c-fibres and impossible to have firing c-fibres without having a headache.⁵ Some physicalists wish to accept the latter impossibility claim while rejecting the former. They hold that mental properties are physical in the sense that mental properties are necessitated by physical properties and necessitate the instantiation of physical properties, but they also hold that the relation between physical and mental properties is many–one, not one–one, as reductive physicalism has it.⁶ Thus, there can be different physical properties besides having firing c-fibres that necessitate pain. This yields a form of physicalism that is weaker than reductive physicalism and hence non-reductive.

We can make non-reductive physicalism more precise by using the notion of strong supervenience, which is defined as follows:

Strong supervenience: A set of properties **A** *strongly supervenes* on a set of properties **B** if and only if, necessarily, if anything instantiates some property *F* in **A** at a given time, then there is a property *G* in **B** such that that thing instantiates *G* at that time, and, necessarily, everything that instantiates *G* at a given time also instantiates *F* at that time.⁷

Here and throughout, ‘necessarily’, unless qualified further, expresses metaphysical necessity, that is, truth in all possible worlds or, somewhat more informatively, truth come what may: if pigs were to fly, donkeys were to talk, and particles were to travel faster than light, what is metaphysically necessary would still have been the case. The definition of strong supervenience is a bit cumbersome, but the underlying idea is simple. As an illustration of the definition, consider dot-matrix pictures and their symmetry properties.⁸ The symmetry properties of a dot-matrix picture strongly supervene on the arrangement of dots in the picture’s matrix. According to the definition, this means that, necessarily, if a picture instantiates a symmetry property at a given time, the dots in the picture’s matrix are arranged in a certain way and that this arrangement necessitates the symmetry property whenever

⁵ See Kripke 1980. I assume, with Kripke, that the names for mental and physical properties are rigid designators, that is, that they name the same property at every possible world. Lewis, by contrast, takes the names of mental properties to be non-rigid designators, which opens up the possibility of contingent psychophysical identities (see Lewis 1980).

⁶ Alternatively, we could characterize the position as saying that mental properties are identical to *higher-order* physical properties, which are distinct from, but stand in the necessitation relation to, sufficiently *fundamental* physical properties. For further discussion of this strategy, see Pauen 2002.

⁷ My formulation of the definition follows what McLaughlin calls ‘Modal-Operator Strong Supervenience’ (1995: 95). For further discussion, see Kim 1984 and McLaughlin 1995.

⁸ This kind of example is due to Lewis 1986c: 14.

a picture has it. For instance, a 3×3 dot-matrix picture that is point-symmetric has to have the dot arrangement $\begin{smallmatrix} \cdot & & \cdot \\ & \cdot & \\ \cdot & & \cdot \end{smallmatrix}$ or $\begin{smallmatrix} \cdot & & \cdot \\ & & \\ \cdot & & \cdot \end{smallmatrix}$, etc., and any picture that has the arrangement $\begin{smallmatrix} \cdot & & \cdot \\ & \cdot & \\ \cdot & & \cdot \end{smallmatrix}$ has to be point-symmetric, any picture that has the arrangement $\begin{smallmatrix} \cdot & & \cdot \\ & & \\ \cdot & & \cdot \end{smallmatrix}$ has to be point-symmetric, etc. More generally, strong supervenience says that a supervenient property has to be accompanied by some subvening property (that is, by a **B**-property), which in turn necessitates the supervenient property whenever it is instantiated.

If we combine the claim that mental properties strongly supervene on physical properties with the claim that mental properties are distinct from physical properties, we get a version of non-reductive physicalism. Generally, non-reductive physicalists claim that mental properties are distinct from physical properties but maintain that mental properties stand in a relation of metaphysical dependence to physical properties (see Baker 2009). The canonical way of spelling out this notion of metaphysical dependence is to read it as strong supervenience (see Kim 1993). For the purposes of this book, I shall identify non-reductive physicalism with the combination of the distinctness claim and the strong supervenience claim. (I do this mainly because it yields a clear-cut terminology, but nothing hinges on it; alternatively, one could use a different label, say ‘strong supervenience physicalism’, for the position thus characterized.) Thus, we get the following definition:

Non-reductive physicalism: Each mental property is distinct from all physical properties, but mental properties strongly supervene on physical properties.

According to non-reductive physicalism, the property of having a headache is distinct from the property of having firing c-fibres and from all other physical properties. But, owing to the strong supervenience of mental properties on physical properties, it is impossible for someone to have a headache without instantiating some physical property that in turn necessitates having a headache.⁹ In my case, that physical property is the property of having firing c-fibres, but in other cases (actual or merely possible), it might be the property of having firing x-fibres (which, let us assume, are actually present not in humans but merely in octopuses), the property of having an active semiconductor network of a certain kind in

⁹ It seems that this necessitation relation holds only if the relevant physical properties include properties to the effect that certain background conditions obtain and that certain laws of nature hold. This issue will be taken up again in Section 4.4.

one's head, etc.¹⁰ Thus, non-reductive physicalism allows for the multiple realizability of mental properties by physical properties.

Some theorists reject not only that mental properties are identical to physical properties, but also that they strongly supervene on physical properties. The view they advocate is dualism, defined as follows:

Dualism: Each mental property is distinct from all physical properties, and no subset of mental properties strongly supervenes on physical properties.

According to dualism, the property of having a headache is not merely distinct from all physical properties, but can be instantiated without a physical property that would in turn necessitate the property of having a headache. In my case, having a headache is accompanied by having firing c-fibres, but, according to dualism, it is possible for there to be someone with firing c-fibres who does not have a headache. Likewise, according to dualism, it is possible in principle for someone to have a headache without being in any physical state whatsoever.

Dualism can take more or less extreme forms. An extreme form might have it that in some remote corner of the universe there are disembodied creatures with headaches, and that, next year, humans with firing c-fibres will no longer have headaches while erupting geysers will have headaches. Few scientifically minded people accept such an extreme form of dualism. According to a more moderate form of dualism, mental properties are tied to physical properties by laws of nature in a way that is structurally similar to, but modally weaker than, strong supervenience (see Chalmers 1996: 123–171). We can express this moderate form of dualism more precisely by introducing the notion of nomological supervenience:

Nomological supervenience: A set of properties **A** *nomologically supervenes* on a set of properties **B** if and only if it is nomologically necessary that if anything instantiates some property *F* in **A** at a given time, then there is a property *G* in **B** such that that thing instantiates *G* at that time, and it is nomologically necessary that everything that instantiates *G* at a given time also instantiates *F* at that time.

The definition of nomological supervenience is just like the definition of strong supervenience, except that the two occurrences of 'necessarily' (which, remember, we stipulated to mean metaphysical necessity) are

¹⁰ The doctrine of non-reductive physicalism is often associated with the view that particular, token mental events are identical to particular physical events. We shall see in the following section, however, that not all accounts of token events allow this. For further discussion of the relation between non-reductive physicalism and token identity, see Schneider 2012.

replaced by ‘it is nomologically necessary that’.¹¹ Nomological necessity is necessity in view of the laws of nature: something is nomologically necessary just in case it is strictly implied by the actual laws of nature. (Equivalently, something is nomologically necessary just in case it is true in all possible worlds in which all actual laws of nature hold.) We can use the notion of nomological supervenience to formulate the moderate form of dualism, which, following Chalmers (1996), I will call naturalistic dualism:

Naturalistic dualism: Each mental property is distinct from all physical properties. No subset of mental properties strongly supervenes on physical properties, but mental properties nomologically supervene on physical properties.

Naturalistic dualism is obviously a version of dualism as we have defined it. Both reductive physicalism and non-reductive physicalism are versions of physicalism about mental properties, but do not exhaust it. One could hold that some but not all mental properties are identical to physical properties, while the remaining mental properties strongly supervene on physical properties. Then one would neither be a reductive physicalist nor a non-reductive physicalist. The best way to describe such a view is to divide the subject-matter and say that its adherents are reductive physicalists about those mental properties that they take to be identical to physical properties, but non-reductive physicalists about those mental properties that they take to merely strongly supervene on physical properties. Similarly, dualism and physicalism do not exhaust logical space. One could hold that some subset of mental properties strongly supervenes on physical properties (perhaps even that the members of this subset are identical to physical properties), while some other subset of mental properties does not strongly supervene on physical properties. Then one would be neither a physicalist nor a dualist. This kind of view seems odder than the version of physicalism that is neither reductive nor non-reductive, but one could again say that its adherents are physicalists about the first subset of mental properties and dualists about the second.

¹¹ Indeed, if it weren’t for the stipulation that ‘necessarily’ mean metaphysical necessity, nomological supervenience would be a kind of strong supervenience, for without the stipulation, strong supervenience could accommodate different kinds of necessity, including nomological necessity. If one preferred this more flexible notion of strong supervenience, one could define two species of this genus, say ‘metaphysical strong supervenience’ and ‘nomological strong supervenience’, which would correspond to our notions of strong supervenience and nomological supervenience, respectively. Our terminology has the advantage of brevity, however.

One could avoid the possibility of hybrid views by defining the positions differently. In particular, one could define non-reductive physicalism as the strong supervenience thesis conjoined with the claim that *some* mental properties are distinct from physical properties. Similarly, one could define dualism as the claim that *some* mental properties are distinct from physical properties conjoined with the claim that *some* subset of mental properties does not even strongly supervene on physical properties. The resulting positions would exhaust logical space, and they would be entailed by, but not entail, the corresponding positions according to the definitions I have given.¹² It seems to me that the stronger, albeit non-exhaustive, definitions capture the general usage better, but one could make a different terminological choice.

Setting aside the issue of whether or not the definitions should exhaust logical space, there is also some controversy about whether the labels ‘reductive physicalism’, ‘non-reductive physicalism’ and ‘dualism’ are apt for the positions to which I have attached them. Some would object to the definition of reductive physicalism because they think that reduction requires more than identity.¹³ Some would object to the definition of non-reductive physicalism because they think that physicalism (non-reductive or otherwise) requires more than strong supervenience.¹⁴ Some would object to the definition of naturalistic dualism because they think that nomological supervenience is sufficient for physicalism and hence for the falsity of dualism (see Kim 2005: 49). (Consequently, they would also object to the definition of dualism *simpliciter*, since, according to the definition, naturalistic dualism is a species of dualism.) I will not address these objections here. It suffices for our purposes that the positions that I have labelled ‘reductive physicalism’, ‘non-reductive physicalism’ and ‘dualism’ have been sufficiently prominent in the philosophy of mind. It does not matter for our purposes whether they really deserve these labels.

There are certain standard objections to each of the positions about the nature of mind. There is one class of objections that have nothing to do with mental causation. Non-reductive physicalists argue that reductive

¹² Strictly speaking, the entailments hold only on the assumption that there are mental properties, which I take for granted here.

¹³ See van Riel 2013. Kim (2005: 34) defines non-reductive physicalism (*inter alia*) as the conjunction of the claim that mental properties are not reducible to physical properties and the claim that they are not identical to physical properties. He explicitly acknowledges that identity is necessary for reduction, however (2005: 34), and his arguments for reductive physicalism (2005: 32–69) suggest that he takes it to be sufficient for reduction too. For discussion of the notion of reduction in the context of emergentism, see Stephan 2002.

¹⁴ See Wilson 2005; see also Jackson 1998: 22–23 and Melnyk 2003: 49–70.

physicalists cannot explain the multiple realizability of mental properties, such as the fact that headaches can be accompanied by c-fibre firings as well as x-fibre firings. Dualists argue that reductive physicalists and non-reductive physicalists cannot explain the possibility of zombies, that is, beings that are physically exactly like us but without any conscious thoughts. I am not going to discuss this class of objections. The comparative merits and difficulties of the different views about the nature of mind will occupy us only where mental causation is concerned.

1.3 The Relata of Causation

Statements of causation come in many stripes. We say that my throwing the stone caused the shattering of the bottle (more idiomatically: that it caused the bottle to shatter); that the reef caused the leakage; that smoking causes cancer. When we say that the throwing caused the shattering, we are talking about token events, that is, particular events (the throwing and the shattering). When we say that the reef caused the leakage, we are talking about a particular thing (the reef) and a particular event (the leakage). When we say that smoking causes cancer, we are talking about general phenomena (smoking, cancer); these might in turn be event types or properties. Perhaps these different statements of causation talk about different kinds of causation that are mutually irreducible; perhaps they do not.¹⁵ In any event, the problems of mental causation are primarily problems about causation between particular events, so my focus will be on this. The interaction problem is about how, in principle, particular mental events such as my headache can cause particular physical events such as my hand's moving towards the aspirin. The exclusion problem is about how a particular physical event such as my hand's moving towards the aspirin can have both a particular mental event and a simultaneous particular physical event as its cause without being like a case of overdetermination.¹⁶ This is not to say that properties play no role in these causal relations – indeed, we shall see that they play a crucial role – but the causal relata are best taken to be particular events. Henceforth, when talking about events without further qualification, I shall mean particular events.

¹⁵ One promising approach analyses claims such as 'Smoking causes cancer' as generic statements about token events. See Carroll 1988, 1991 and Swanson 2012b for further discussion.

¹⁶ Lowe (2000, 2008: 41–57) advocates a solution to the exclusion problem according to which mental events cause not physical *events* but *facts* about intra-physical causal relations. If this kind of solution could be made to work, it would still be only a second-best solution. It would be better to have a solution in terms of causation between events.

What are events? In particular, when are we dealing with a single event and when with several events? W.V.O. Quine and others think that events are identical just in case they occur in the same spatiotemporal region.¹⁷ This account individuates events in a rather coarse-grained way. I stroll leisurely. By virtue of strolling leisurely, I stroll. My strolling and my strolling leisurely take place in the same spatiotemporal region. Therefore, according to the Quinean account, my strolling and my strolling leisurely are one and the same event. Someone who is worried about multiplying events beyond necessity will welcome this result. But there are problems. Perhaps my strolling leisurely, but not my strolling *per se*, causes me to feel refreshed afterwards. How can my strolling leisurely cause something that my strolling does not cause, yet be the very same event as my strolling? By Leibniz's law, identical events must have the same effects (and the same causes). Other examples bring out this issue more sharply. A metal sphere rotates and heats up at the same time. (The heating is due to an external source.)¹⁸ The sphere's rotating and the sphere's heating up take place in the same spatiotemporal region. Therefore, according to the Quinean account, the sphere's rotating and the sphere's heating up are one and the same event. This should sound strange even to those who are inclined towards ontological parsimony. And to say that the sphere's rotating and the sphere's heating up have all their causes and effects in common sounds even less plausible than the parallel claim in the case of my stroll. If I place a funny hat on the sphere, it will start rotating too, but the hat's rotation will be caused by the sphere's rotating, not by the sphere's heating up.

These problems force us, I think, to reject the Quinean account. This may seem unfortunate, for the account seems to offer an attractive solution to the problems of mental causation. Suppose that my headache takes place in the same spatiotemporal region as my c-fibre firing does. (The claim that headaches have spatial location is not completely uncontroversial, but anyone who is a substance physicalist should find it acceptable.) Then, according to the Quinean account, my headache and my c-fibre firing are one and the same event.¹⁹ Thus, my headache and my c-fibre firing have all their causes and effects in common. The interaction problem

¹⁷ See Lemmon 1967, Smart 1972, Quine 1970: 31–32, Quine 1974: 5, 131–132, and Davidson 1985. Previously, Davidson (1969) had endorsed the individuation of events in terms of sameness of causes and effects, which critics had claimed to be covertly circular.

¹⁸ The example is from Davidson 1969.

¹⁹ This result is a token identity claim, albeit one that, if generalized, remains weaker than Davidson's (1970) famous Anomalous Monism, because it does not claim that the mental is anomalous.

disappears. No one denies that my c-fibre firing can cause my hand to move towards the aspirin; if it does, then *ipso facto* my headache causes my hand to move towards the aspirin. The exclusion problem disappears, too. Since my headache and my c-fibre firing are one and the same cause of my hand's moving towards the aspirin, my hand's moving is not caused twice over. Even non-reductive physicalists and dualists about mental properties could invoke this solution to problems of mental causation, for what matters is merely that particular mental and physical events are identical; the mental and physical properties that are involved can be distinct and need not even stand in a relation of strong supervenience.

Apart from requiring an implausible conception of events, however, this solution to the problems of mental causation loses its appeal on closer scrutiny. When we demand an explanation of how physical effects can have mental causes, we want to know how mental events can cause physical events by virtue of their mental properties. The explanation that has been suggested provides at most an account of how mental events cause physical events by virtue of their physical properties. According to the explanation, my headache causes my hand to move towards the aspirin because it is identical to my c-fibre firing. Presumably, how my c-fibre firing causes my hand to move is in turn explained by the physical properties involved in my c-fibre firing. So we are still lacking an explanation of how mental events can be causes *qua* mental. For all we know, the situation might be like the following: I put an apple on a scale. The apple weighs 100 grams and has a temperature of 20 degrees Celsius. Shortly after putting the apple on the scale, the display flashes '100'.²⁰ Quineans have to say that the apple's weighing 100 grams (at a certain time shortly before the flashing) is identical to the apple's having a temperature of 20 degrees (at that time), since these events take place in the same spatiotemporal region.²¹ We may assume that the apple's weighing 100 grams caused the display to flash '100'. If the apple's weighing 100 grams is identical to the apple's having a temperature of 20 degrees, it follows that the apple's having a temperature of 20 degrees caused the display to flash '100'. This is an implausible result similar to the result that the sphere's heating up caused the hat to turn, but let us accept it for the sake of argument. The temperature causes the flashing, then, but we have not yet explained how it causes the flashing *qua* temperature.

²⁰ This example is a variation of an example from Honderich 1982.

²¹ I am assuming here and throughout that events can be static, that is, that events need not involve change.