

# *Physicalism and the Mental: The Dominant View*

Will's perception of his father coming out of the cinema with a lover caused Will's immediate upset and his subsequent behaviour. This is how we would naturally describe what happened. From this perspective, Will's mental contents appear as distinctly active, as making a difference to what occurs next. Variations in Will's perceptions would have given rise to changes in his emotional reactions and behaviour. We take it that, for instance, if Will had not seen Hannah Plowman as his father's lover, there would not have been any interesting story to report. Still, this mentalistic narration appears to call for a complement. It seems that a neurophysiological sequence must somehow accompany Will's perceptual process; after all, if Will is to have some perceptions, be upset or move his body, significant changes must take place in his physiology. The problem that concerns us is to find a credible account of how exactly these two processes, namely, the neurophysiological and the mental, are supposed to interrelate. What do our metaphysical intuitions tell us in this respect? This question articulates the contemporary debate about mental causation.

To draw a precise map of the available answers to this question is a hard job, since pressure in the academic environment encourages people to adopt every possible position on the issue, and the scenario becomes as intricate as a spider's web. Still, the range of possible stances seems constrained by the presuppositions one has to grant in order to be recognized as an intelligible voice in the debate. And, in this respect, to treat a certain point of view as privileged, as the position with regard to which any other perspective is to be located, is part of what one should initially grant to enter a philosophical controversy. We consequently propose to introduce the current discussion on the causal efficacy of the mental by

putting forward, in Chapter 1, what we take it to be that dominant perspective on the issue.

As we see it, the dominant approach represents an attempt to articulate two seemingly fundamental metaphysical intuitions, namely: the causal efficacy of the mental and the causal primacy of the physical. Moreover, they assume that our intuitions about the causal efficacy of the mental could be properly honoured only if one could show how our mental states are causally efficacious in virtue of their mental properties. So, the crucial question becomes how mental properties could be causally relevant while, nevertheless, respecting the causal primacy of physical properties. '*Causal physicalism*' is the name we propose to designate the kind of answer that the dominant perspective provides for that question and, therefore, it is meant to pick out a rather general metaphysical picture, whose details may be filled in several different ways. In this sense, '*causal physicalism*' alludes to a set of theories that share a certain assumption, namely: that every case of nonphysical (and hence mental) causation has to be conceived as systematically dependent on certain physical processes that underlie it. In other words, the dominant view assumes that the only way to render intelligible the idea of non-physical causation is by reducing its autonomy, at least in this minimal sense. Whenever a genuine cause produces an effect, the causal efficacy of the nonphysical properties involved in that process will ultimately depend on the instantiation of certain basic (physical) properties. For whatever the causal efficacy of a nonphysical property may be *in a particular case*, it cannot go beyond the causal powers of the physical properties that are involved *on that particular occasion*.<sup>1</sup> Hence, physical properties appear to the causal physicalist as the basic properties of the world, while the rest of the properties are realized in terms of those putative basic properties. The *dominant view* may then be provisionally characterized as the attempt to vindicate the causal efficacy of the mental by showing that mental properties meet the demands that causal physicalism impose.

There is, indeed, an important problem that such a view has to face, namely: to specify what those basic, physical properties are. Yet, we intend to be rather concessive with our opponents on this issue. It is true that some of the problems we will raise against causal physicalists in later chapters will turn out to be crucially related to their difficulties to provide a coherent account of the metaphysical virtues attributed to those physical properties. But, in any case, as we should talk in this chapter about the metaphysical requirements that physical and non-physical properties are supposed to satisfy, we will not impugn the content of those principles on the basis that they would require a previ-

ous specification of what a physical property may be. Instead, we will take for granted that there might be a certain minimal and interesting subset of properties that could satisfy such requirements. This concession fits the overall character of our challenge to causal physicalism, which has the structure of a *reductio ad absurdum*. Even granting certain metaphysical principles that are proposed to motivate causal physicalism, this metaphysical position is in trouble mainly because they are forced to accept certain conclusions (for instance, massive overdetermination) that no one (least of all the causal physicalist) is prepared to concede.

So, we can say that the main purpose of this book is to elaborate a number of *reductio* arguments that are meant to challenge causal physicalism, as well as the dominant view, in its attempt to show the compatibility between the causal efficacy of the mental and causal physicalism. As a result, we will provide an alternative analysis of our intuitions about the primacy of the physical that fits with the causal autonomy of mental properties.

In this context, we shall devote the present chapter to characterize and motivate the dominant view. In section 1, we intend to locate the dominant stance within the debate on mental causation and, in section 2, we begin to describe the metaphysical motivation for causal physicalism. Thus, a physicalist principle is introduced which is meant to express a supposedly elementary intuition concerning the causal closure of the physical world, as well as a statement that denies the existence of more than one complete and independent explanation of a single effect. These two considerations jointly motivate, in section 3, a physicalist constraint that every causally responsible property must comply with. 'Strong Supervenience' is the tag attached to this physicalist requirement, and heralded as one of the chief convictions of the dominant stance.<sup>2</sup> In section 4, we elucidate how physical properties must be conceived if physicalist intuitions have to make sense; secondly, we connect Strong Supervenience with certain images about the intrinsicness of causally efficacious properties; and, finally, we emphasize that Strong Supervenience falls short of what the dominant view should demand and a more stringent physicalist constraint is proposed instead. To close, we explore the reasons why some people think that the causal efficacy of content may be at odds with the physicalist constraints stipulated by the dominant stance. From the idea that mental causation has to be conceived as supervenient on the physical properties of certain objects, the dominant view tends to derive the claim that content can be causally efficacious only if it is narrow – only if it is not fixed by relational properties. A substantial portion of our labour in Chapters 2–4 will be to show that this

line of reasoning may have devastating effects, since it not only poses a problem to mental content, but also to functional properties and, in general, to the properties posited by special sciences such as biology, geology, and so forth. These severe results will prompt a revision of the arguments in favour of causal physicalism and, therefore, of the conditions that causally efficacious properties are due to meet.

## **1 The Dominant View in the Debate on Mental Causation**

If it is really true that the content of Will's perception had a certain effect on some of his other mental states, we should accept that certain differences in the content of his perception would have produced some differences in his other mental states. And, of course, the same is true regarding the causal efficacy of his mind on his behaviour. In fact, we accept that some aspects of that behaviour are an effect of his becoming upset by the perception of his father's lover only because we assume that, had he felt happiness when realizing that this particular lady was his father's lover, his reaction would have been rather different. In general, we can then say that a mental state has certain effects in virtue of the properties of such a state, since differences in the properties of a mental state alter its capacity to bring about certain effects.

There are, indeed, three ways in which mental properties must be able to participate in causal processes. One, minds are affected by what happens in the physical world: the content of Will's perception seems to be causally dependent on certain inputs that reach his open eyes and alter his nervous system. Two, mental states can also influence other mental states: Will's perception of his father and the lady produces certain beliefs that disturb him. And, three, mental states can be causally responsible for our way of acting, as well as for some physiological changes in our body: Will's mood is the cause of his rushing towards home, and of the strain in his stomach too.

Unsurprisingly, an important part of the philosophical literature on mental causation has focused on the first and third cases, where the mental is claimed to interact with the physical. For most of the hardships about mental causation derive from the fact that minds do causally interact with the material world. It seems that if we were able to understand this interaction, we would also be able to account for mental-to-mental processes, and mental causation would stop being a problem.

How is it, then, possible for our minds to move the material world? How is it possible that certain mental properties could be causally

responsible for certain physical changes in the world, or vice versa? As one should expect, the dominant answer to this question emerges out of the rejection of a number of alternatives. We shall, thus, explore in this section the reasons why those alternatives look scarcely attractive to the contemporary eye. First, we shall object to substantial dualism, and then discount type-identity theories insofar as they may fail to respect the multiple realizability of mental properties. This will give rise to functionalism, which represents the version of causal physicalism to which the dominant view subscribes. We shall see, however, that functionalism may not be so different from type-identity theories as one might expect, and that this similarity may point to a problem for the causal relevance of functional properties. We close by emphasizing how mere token-identity theories cannot solve the problems that functional theories may encounter in their attempt to account for the causal efficacy of the mental.

*Substantial dualism* is almost universally envisaged as a manifestly unsatisfactory answer to the question as to how mental properties may intervene in the physical world. A classical defence of dualism can be found in Descartes' doctrines about mind and body interacting through the pineal gland. Descartes' idea was that *res cogitans* and *res extensa* should be conceived as two different kinds of stuff, that mental and physical properties are determinations of those different stuffs. The question is: how can determinations of such different substances interact? Descartes' own way of explaining this interaction shows how implausible his solution was. As some of his contemporaries realized, it is not easy to accept that a substance without extension could causally intermingle with the physical world.<sup>3</sup> This problem cannot be tackled by simply fixing a privileged point in the physical world, e.g. the pineal gland, where both substances actually interact. This kind of solution sounds as implausible as a hypothetical defence of the causal activity of the soul by a modern scientist who chooses to appeal to the (supposedly) empirical fact that soul and body interact in the brain. In fact, it appears that no empirical discovery could really inform us about the putative ability of an immaterial substance to move the material world.

Nowadays these critical intuitions tend to be expressed along the following lines. Causation cannot be a matter of magic; therefore, it should be taken for granted that, for any causal chain that affects the physical world, there must be a physical process that can explain the effect. The relevant intuition can be expressed as an intuition regarding explanation or as an intuition regarding causation. Let us turn again to Gordimer's novel. Will is upset because of a certain perceptual experience. His being upset is the cause of a certain piece of behaviour. Nevertheless, it seems

that we can assume that there is a complete causal explanation of his movements that does not include any mental predicates. This explanation might mention, for instance, electrical changes in his brain, neurophysiological processes, movements in his muscles, etc. Following the current dominant tendency in the philosophy of mind, we may refer to this kind of explanation as a 'physical explanation' and, therefore, escape by now any controversy about the status of chemical or neurophysiological predicates with regard to physics, that is, to the most basic science of nature. What seems to be clear, in any case, is that chemical and neurophysiological predicates could not ever count as mental.

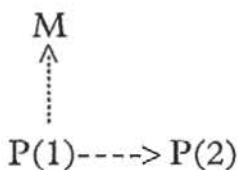
These intuitions can also be expressed in metaphysical terms: *prima facie*, it seems that there is a chain of physical causation that is enough to produce the physical effect and that does not require the instantiation of any mental property. Perhaps we do not know the entire physical story, but we assume that it must exist. This does not amount to saying that at some time scientists will actually be able to provide such a story. The point being raised is just a matter of principle: the dominant view assumes that, if the causal story they told us were not a complete one, a completion of it would require additional statements in which no mental term could properly appear. Every cellular reaction in our body demands an explanation which has to refer to the kind of process that we accept as a plausible candidate to excite cellular activity – and, certainly, it is not here where mental properties can find their proper causal role.

There are not many moves that a dualist can make to face this basic intuition. And none of them has the slightest possibility of succeeding. One possibility could be to conceive the situation as a case of overdetermination, like the death of a man who simultaneously receives several lethal stabbings. In each case in which our mind causally affects our body, there would at least be two different causal chains involved in the production of the same effect, both of them sufficient for it but none of them necessary. On this view, mental properties would not intervene in certain causal (physical) chains that produce a given effect, but they can intervene in some other parallel causal chain producing the very same effect. So, whenever our mind has causal influence on our body, the different causal chains fixing the relevant outcome are not necessary for it. Yet, it is commonly assumed that massive overdetermination counts as a *reductio* of any theory that entails it. For causes are counterfactually necessary for their effects.

There is no advantage in retouching the image and conceiving the mental chain as only 'helping' (i.e. not being enough) for the relevant physical outcome. Under the assumption that the physical chain is

enough, the mental contribution comes up as irrelevant. Had the mental property been absent the physical chain would still have done the causal job. Consequently, it can never be true of Will that, if he had not seen his father's lover and changed his mood because of this, he would not have reacted the way he did. So, it does not seem that the dualist has apprehended the idea that our minds can be causally responsible for certain effects.

There is a different way in which a dualist may try to deal with this issue, namely: to conceive changes in the mental substance as epiphenomena, as effects of certain causal processes that have no causal consequences. By this move, the dualist seeks to preserve the common intuition that mental states are necessary for certain events, while avoiding a conflict with the idea that there is always a sufficient causal, physical explanation. Yet, the price to be paid is denying the causal efficacy of mental properties. John's desire to drink a beer could be the effect of certain neurophysiological processes in John's brain, and those processes might also be the cause of certain movements of John's body. But the mental state itself would be deprived of any power to alter John's body. It is true, though, that even in this context, it could be argued that John's mental states are a necessary condition for its bodily movements. For a mental state *M* could be conceived as an inexorable outcome of a neurophysiological chain *P*(1) that is, in turn, a necessary causal condition for a physical effect *P*(2).



One could thus say that, if *M* had not been instantiated, *P*(2) would not have been instantiated either. For, without *M*, *P*(1) could not have been instantiated, given that *P*(1) is supposed to be a sufficient condition for *M*. Then, *P*(2) would not have been instantiated either, given that *P*(1) is supposed to be a necessary condition for the instantiation of *P*(2). Yet, if these are all the causal powers that mental properties may gather, we may safely affirm that mental properties have no causal power at all. Mental 'properties' become mere epiphenomena, and both philosophers and nature dislike epiphenomena, since causal efficacy counts as a mark of reality.

All this has some crucial implications for the contemporary debate on mental causation. For the previous reflections not only supply an argument against substantial dualism, but also involve an explicit

demand for some kind of connection between mental properties and the physical world. Furthermore, given that the main objection to substantial dualism has been that it concedes no relevant causal role to mental properties, it seems pointless to deny the existence of ghostly minds without saying anything about how the causal efficacy of mental properties is related to the causal efficacy of the nonmental chain that, we are assuming, is enough to move our bodies.

There is indeed quite a natural way of dealing with these difficulties. If mental properties were identical to some of the features of the material causal chain that is enough to move our body, then mental properties could after all be vindicated. This is the basic line of reasoning behind the so-called *type-identity* theories.<sup>4</sup> In fact, the previous criticism of dualism crucially depends on the reluctance to accept any causal role for minds if there are complete physical explanations of their putative physical effects. But this objection suggests the idea that, no matter what the causal powers of mental states may be, they must be identical with the causal powers of certain subsets of the physical (neurophysiological) states that are supposed to be enough to fix the complete physical (neurophysiological) explanation. And so we reach type-identity theories.

Nevertheless, there is an important problem that has put type-identity theories out of fashion for many years, namely: the multiple realizability of mental properties. For, there is no reason to think that a mental type (for instance 'to be in pain' or 'to believe that Paris is in France') must have the same type of neurophysiological instantiation in any possible mind. If that were so, we could never be justified in attributing mental states to any creature unless we were confident that its neurophysiological make-up was identical to ours. And that seems an excessive requirement, since we are prepared to consider the possibility of creatures that have a mind despite their respective biology being quite different from ours.

This difficulty is decisive to understand the appeal of *functionalism* in the contemporary philosophy of mind. Functionalism is, undoubtedly, the philosophical conception of mind that nowadays prevails. An essential feature of functionalist accounts of mind is to assimilate mental properties to functional properties and, by so doing, individuate them in terms of their causal role. This causal role is fixed by the functional theory that defines the property at stake. This is best understood by considering, first, a nonmental case, say a brake: to be a brake is to have certain properties allowing a particular object to play the causal role that is fixed by a particular definition of 'brake'. Consider, for instance, the following one: 'to be an object such that, when pressed, triggers the



action of a mechanism that, by exercising certain force on the wheels, reduces the speed of a moving vehicle'. In the same way, it is supposed by functionalists that 'to be in pain', 'to have the desire of eating an apple', or 'to believe that Paris is the capital of France' can be defined by appealing to their typical effects on our behaviour, given certain other beliefs and desires.

We should notice two apparent virtues of functionalism. At first glance, the causal efficacy of brakes does not seem to be in tension with the fact that every physical effect has a physical causal explanation. Thus, if mental properties are vindicated as functional properties, their causal efficacy will be of the same kind as the causal efficacy of brakes: this causal efficacy does not seem, at least *prima facie*, to be jeopardized by the existence of an implementing physical mechanism that is enough to produce the effect. Secondly, functionalism is a conception of the mind specially designed to deal with the multiple realizability of non-basic properties, and, therefore, has no trouble accounting for this fact in the case of mental properties. An object made of iron can satisfy the requirements that a functional theory imposes on the property 'being a brake', but brakes can be made of many different kinds of material. Similarly, a functional theory of mind leaves open the possibility of a mental property being realized in many different ways.

The notion of 'realization' is crucial to our discussion. On a rather conventional understanding of functionalism, a mental state is a second-order state: the state of having some other properties (say, neurophysiological properties in the case of the human species) that play a certain causal role. These properties constitute in each particular case the realization of that mental state. Does this approach involve a commitment to what we have called 'causal physicalism'? Functionalism by itself does not require the physical or neurophysiological realization of mental properties. However, if functionalism is attractive with regard to the causal efficacy of the mental, it is just because it is connected with the principle that mental states require material realizations. It is only after this connection that we can see functionalism as a particular version – the predominant one nowadays – of causal physicalism. Thus, from the viewpoint of functionalism, my being in pain here and now is fixed by my instantiating here and now certain neurophysiological or physical properties. Functionalism takes it that in any particular case in which a mental property is causally efficacious there is a physical realization of this property that is causally active. Whatever the causal efficacy of the mental property may be *on a particular occasion*, it cannot go beyond the causal powers of the physical or neurophysiological properties that realize it *in that particular case*. Functionalism (hereafter, construed as

including the demand that functional properties should be physically realized) appears then as a special version of causal physicalism, namely, the version that inspires the dominant view.<sup>5</sup>

Our attack against causal physicalism (and, hence, against functionalism) will be relatively independent of conventional discussions about the best way to characterize the connection between functional properties and their respective realizations. Some aspects of this discussion are merely terminological. According to David Lewis, for instance, functional terms can be used in two different ways. In the first one, they are used as nonrigid designators, so that a functional term refers to a number of different properties, as many as there are different 'realizations'. This sense of functional terms implies that multiple realizability does not constitute an interesting metaphysical phenomenon, since it is just the result of using the same (functional) term to designate different properties (i.e. the so-called 'different realizations') in disparate contexts. From this perspective, a functional property would not differ from what is usually called its 'realization'. So, if mental terms are functional terms, a mental predicate such as 'pain' will refer to different (neurophysiological) properties.

Is there any other property that might be picked out by our functional terms? Is there any genuine property that could be shared by all the realizations of a second-order property? Lewis has granted that, in the sense of a functional term that he calls 'diagonalized', an expression such as 'being in F' might be interpreted as picking out a property Q that would be shared, across all physically possible contexts, by all occupants of the causal role that is proper to 'F'.<sup>6</sup> Nevertheless, Lewis has argued that Q itself cannot be causally efficacious. For, according to him, Q does not have causal powers of its own, apart from those of their different respective realizations: it is so disjunctive that it cannot fix the causal efficacy of any particular event.<sup>7</sup> At this point, his position is quite similar to Jaegwon Kim's. For, according to the latter, a second-order property is to be conceived as an enormous disjunctive set (i.e. the disjunction of all its possible realizations) and consequently, the causal powers of each instantiation of a second-order property are identical with those of its specific realization. Hence, functionalism seems to ultimately coincide with type-identity theories, given that the phenomenon of multiple realizability only shows that the notion of identity has to be relativized to the different types of realization. So, if it happens that a mental property is identical with a certain neurophysiological property among the members of a species, this does not entail that this identity has to be preserved in other species. In fact, it might occur that the iden-

tity has to be relativized to different individuals in the same species, or to different moments in the life of a given individual.<sup>8</sup>

At first sight, those philosophers who envisage functionalism as a specific version of type-identity theories are able to honour the causal efficacy of mental properties, since mental types would just coincide with neurophysiological types. Yet, this approach is connected with a resilient worry today. Several people have argued that functionalism may not be able to preserve the causal efficacy of mental properties precisely because it has to be ultimately construed as a mere version of type-identity theories. This line of reasoning may be sketched as follows. Philosophers were initially fascinated by the idea that the connection between functional properties and their 'realizations' could provide a good model for mental properties, just because (i) the multiple realizability of functional properties, and (ii) the apparently obvious way in which the causal efficacy of functional properties does not pose any problem to a physicalist conception of the world. But, after some decades of functionalism as the dominant paradigm in the philosophy of mind, the autonomous causal efficacy of functional properties does not seem such an obvious thing. To take an example from Ned Block: consider the dormitive power. This is a second-order property: the property of having some first-order property (for instance, a certain chemical composition) that induces sleep. Yet, some may object that it would be very strange indeed to consider that this functional characterization helps at all to vindicate the causal powers of dormativity as a second-order property, since the only causal powers involved here are those fixed by the chemical composition. But what goes for dormativity seems to apply to every functional property. So, it seems that, following Lewis and Kim, we should envisage functionalism as a variant of type-identity theories. But, then, how could we retain the causal powers fixed by functional (and mental) predicates as different from those of their respective realizations? In any event, we do not intend to directly address this issue for now, since our main argument against causal physicalism is independent of the position that a functionalist could take on this controversial question.<sup>9</sup>

A functionalist has to grant causal physicalism anyway and, therefore, recognize that in any particular case in which a mental property is causally efficacious, its causal efficacy is fixed by the causal efficacy of certain physical (neurophysiological) properties that realize it on that particular occasion. According to the dominant view, this is the only way in which our minds may move the physical world. In the rest of this book we will often employ the language of functionalism. The main reason is that, in the present controversy, it surely constitutes the dominant

elaboration of that physicalist stance. Nevertheless, our arguments against functionalism tend to have a wider scope, since they affect every treatment of causal dispositions that is compatible with causal physicalism. For instance, if our argument against functionalism is right, it would trivially follow from it that we cannot give any sense to the idea that a determinate set of neurophysiological properties are identical to a mental property.

Whatever the difficulties that functionalism has to face, we must be aware that functionalism is something more than a mere token-identity thesis. Of course, for a functionalist, the instantiation of a certain mental property, here and now, is just the instantiation, here and now, of a certain set of nonmental (neurophysiological/physical) properties. Hence, functionalism insists that tokens (i.e. instances) of mental properties are identical to tokens (i.e. instances) of physical (neurophysiological) properties. But it still tries to tell us an intelligible story about how mental and neurophysiological types are related: mental types are defined by their having certain connections (they are second-order types) with material/physical/neurophysiological types. In fact, the original attraction of functionalism derives not only from its promising capacity to integrate the phenomenon of multiple realizability, as opposed to standard type-identity theories, but also from the dissatisfaction associated with mere token-identity theories. For one could not escape the impression that a mere token-identity theory falls short of apprehending our intuitions about the causal efficacy of the mental.

A mere token-identity theory claims that (i) any mental particular is a physical particular; (ii) without committing itself to any interesting connection between the properties they exemplify. This way one could vindicate mental causation without worrying about how mental and physical properties might interrelate. Yet, it would be a mistake to think that such a position is an easy, direct way of dissolving the difficulties we have been discussing in the previous pages. For, as the dominant view insists, one cannot safely dispense with the causal relevance of properties. To motivate this claim, let us consider Donald Davidson's treatment of mental causation, which invokes, at least *prima facie*, a mere token-identity theory.

Causal relations are, according to Davidson,<sup>10</sup> ontological affairs that hold between events, regardless of how we might choose to describe them or, relatedly, the properties by which we may decide to pick them out.<sup>11</sup> This means that there is no sense in which the causal efficacy of an event can be attributed to its mental properties; since, on this account, properties are not the relata of causal connections. Davidson defends an extreme nominalistic position: an event's 'properties' are just

whatever determines that certain descriptions are true of it. Davidson also thinks that every mental event is a physical event. His way of arguing this point is commonplace in current literature: (1) mental events are causally related to physical events; (2) causal relations among events are to be governed, under some description, by strict laws; (3) only physical predicates can figure in strict laws; hence, (4) every event admits of a physical description and is, therefore, a physical event.

Prima facie, Davidson's account goes against our intuitions about the causal efficacy of the mental.<sup>12</sup> It seems that only by recognizing that events cause other events by virtue of instantiating certain properties, can our intuitions about causal relevance be at all expressed.<sup>13</sup> Mental events will surely cause other events, but it is not easy to grant, as Davidson demands, that mental properties are to be inert as to the deployment of such processes, since there are important intuitions about causal relevance which push in the opposite direction. As we realized in Will's story, we do not simply want to say that an event A ('The perception of his father coming out of the cinema with a woman') caused another event B ('Will's emotional upset'), but that event A having a certain mental property P caused event B having another mental property Q. One could also say that events A and B had plenty of other properties: Will went through a number of physiological changes, the cinema was a particular cinema, Will's perception included other contents (the colour of Will's father's jacket, the lover wearing glasses, etc.), and so on. But the writer rightly assumes that it is only worth mentioning those properties of the events that are relevant to the causal processes at stake.

Davidson can, of course, always appeal to his nominalistic assumptions: causal connections are ontological affairs, and properties are just fixed by our way of describing events. So, he would insist, the whole question about the causal relevance of mental properties is ill formed. But this move does not dissolve the problem. It seems that the idea that causation only holds between events, no matter how they are actually described, or the properties by which they are picked out, simply alters the way in which the worry has to be expressed: we should now talk in terms of causal explanations. Consider, for instance, the following example:

'Something possessing content, or having meaning, can be a cause without its possessing that content or having that meaning being at all relevant to its causal powers. A soprano's upper-register supplications may shatter a glass, but their meaning is irrelevant to their having this effect. Their effect on the glass would be the same if they meant nothing at all or something

entirely different. . . . If brain structures possessing meaning affect motor output in the way the soprano's acoustic productions affect glass, then the meaning of these neural structures is causally inert. Even if it is there, it doesn't do anything.<sup>14</sup>

Faced with this kind of reflection, we cannot confine ourselves to repeating that properties have no causal import, that there are no such things as causally relevant properties. The crucial issue is that, with regard to the shattering of the glass, there is a pertinent difference between the content of the aria and its pitch. We may express this difference, if we wish, in terms of causal explanation: to explain the shattering, we can rightly appeal to the pitch, but not to the content of the aria. The pitch description is explanatorily relevant, while the content description is not. The question is now: why should the content of a mental state (say, the desire to drink a beer) not be as irrelevant with regard to certain physical movements (say, the action of walking towards the fridge) as the content of the aria with respect to the shattering of the glass?

Davidson certainly tried to address this issue by granting a minimal sense in which physical properties (descriptions) are more basic than other properties. According to him, any nonphysical difference between events requires a physical difference.<sup>15</sup> So, he could argue that, if the event 'Will's perception of his father's lover' had been different in its mental properties, it would have been a different event and, therefore, it would have been different in its physical properties. But, even if this might be the beginning of the story, it cannot be the whole of it. For, on this line of reasoning, we should also say that if the content of the song had been different, the physical properties of the singing would have been different and the content of the song should, correspondingly, be judged relevant to the shattering of the glass. And this is not correct. For we still have the intuition that, in the circumstances of Dretske's example, the following counterfactual is false:

Had the song had a different content, the glass would not have been shattered.

While it is true that if the song had had a much lower pitch, the glass would not have been shattered.

To couch our point in a language keener to Davidson, we could say that our causal-explanatory judgements about the relevance of the content and the pitch in our song's ability-to-shatter glass, seem to presuppose that an event A with a completely different content and the same

pitch would have shattered the glass as well, while an event B with the same content and a much lower pitch would not have shattered it. The important lesson to be learned at this stage is not whether Davidson's approach can actually make sense of the truth of these counterfactuals, which, as Dretske's example suggests, seem to express our intuitions about the distinct causal-explanatory relevance of the properties involved in events A and B. This is an issue that we will extensively treat in Chapter 5. Our main point now is simply that, in order to make sense of our intuitions about the truth-values of those counterfactuals, we need to go beyond the mere acceptance of a token-identity theory. The fact that a certain particular instantiates both a mental property and a physical property cannot by itself vindicate the causal (or causally explanatory) relevance of the former.

Before moving on to put forward the metaphysical motivation for causal physicalism, let's briefly justify a few terminological options. It may be clear by now that it is not a mere accident that, when talking about the causal relevance of mind, most people assume that the crucial issue is the causal relevance of mental properties. What matters is not merely the causal relevance of mental particulars, but the efficacy of those particulars in virtue of the mental properties that they instantiate. In everyday talk, we speak of mental events, mental processes, mental states, mental phenomena, mental occurrences, or mental facts. Nothing in our main argument will hang on choosing any of these expressions. When we should talk about 'events', we will assume a conception of events as articulated particulars: a conception according to which some properties of an event – namely, those that are essential to (or constitutive of) the event – are its causally relevant properties. On this terminological option, the philosophical perplexity about the causal relevance of the mind can surely be expressed by asking about the causal relevance of mental events. And this kind of talk would be entirely compatible with the conviction that the crucial issue is the causal relevance of the mental properties that constitute those events.

By accepting that the causal efficacy of the mind amounts to the causal efficacy of mental properties, we are conceding that causal connections hold in virtue of the tokening of certain properties. We come from seeing that even this minimal assumption is not uncontroversial, since Davidson has linked his conception of events as unarticulated particulars to a conception of causation and properties in which no question can be meaningfully asked about the causal relevance of properties. Yet we have provided some motivation to retain that assumption, at least in the context of our challenge to the dominant view. First, because most philosophers who enter the contemporary debate on mental causation

regard the causal efficacy of mental properties as the crucial issue and, consequently, we can claim that the Davidsonian conception of properties is not part of the dominant view. Secondly, we have seen that no interesting issue is answered by shifting from the causal relevance of properties to the explanatory relevance of certain descriptions of events. And last but not least: we hope that the connection between our criticism to the dominant view and the Davidsonian conception of causation and properties should become clear in later chapters of this book. For, if our arguments in those chapters are correct, causation cannot be a connection between unstructured particulars fully independent of the properties that those particulars do instantiate. But, for the moment, let us focus on the metaphysical motivation for causal physicalism and, in relation, for the demand that functional properties should have a physical realization.

## 2 The Causal Closure of the Physical World and Explanatory Exclusion

In an initial reflection about what our metaphysical images are, everyone may be strongly inclined to concede that a physical mechanism must underlie any change that the world actually undergoes. We saw how this image emerged as we raised the case against substantial dualism, or as we couldn't help asking for the physiological mechanisms that intervened in Will's perceptions, thoughts and actions. Had some physical mechanism not been present, nothing would have happened at all. It seems that we cannot accept the existence of a change in the world without the corresponding changes in the physical properties of it.

But this is just an image. We need to define it in more refined terms. Specifically, what is the precise content that the dominant view attaches to this image? A proper answer will require some elaboration, and the job will only be accomplished by the end of section 4. Let us propose the following principle as a start:

**Physical Closure:** Every physical event has a complete causal explanation in physical terms.

This principle is often claimed to apprehend a central portion of our metaphysical intuitions concerning the relevance of the physical.<sup>16</sup> Suppose, now, that there occurs for John an event *c* with a certain content property *M*, and *c*'s having property *M* causes an event *e* having property *B*.<sup>17</sup> The physical closure of the world implies that there



must also be an event  $c^*$  with a neurophysiological property N which causes event  $e$  having property B. For the sake of simplicity, we can safely assume that  $c = c^*$ , since this will not affect the discussion. We have, then, two causal chains:

- Case I:** (1)  $c$ 's having M causes  $e$ 's having B  
 (2)  $c$ 's having N causes  $e$ 's having B

Thus, all that Physical Closure demands is that, for each causal chain of class (1), there must be a causal chain of class (2). We should now consider how these two causal chains interrelate. Are we prepared to concede that  $e$ 's having B may have two alternative causal explanations? It is true that different partial causes of the same effect are often mentioned. The problem is, however,

**Explanatory Exclusion**, i.e. that '... Two or more complete and independent explanations of the same event or phenomenon cannot coexist'.<sup>18</sup>

Or, to put it in another way, 'No event can have more than one complete and independent explanation.'<sup>19</sup> Some seeming counterexamples to this principle are not actually so because, as one easily realizes on second thought, they do not really deal with either complete or entirely independent explanations of a single explanandum.<sup>20</sup> For example, we could say that Will's going to the cinema that afternoon caused his emotional upset, since if Will had not been to the cinema on such an occasion, he would not have been distressed. Nevertheless, this new explanation is both partial and interdependent with the previous explanation, which mentions his meeting his father with a lover. Will's going to the cinema is not a sufficient condition for his getting emotionally upset. Thus, we see that two accounts of a single event may pick out two distinct causes. But, in such cases, both causes will probably be just two partial or interdependent causes of the event, even if each explanation singles out one of them as the most salient.

There is, however, another kind of counterexample to Explanatory Exclusion, namely, those cases in which it could be rightly argued that two different, complete and independent causal chains do overdetermine their effect: two bullets killing a man at the very same time, two darts simultaneously exploding a balloon. But what should we say about these cases? Does overdetermination actually conflict with Explanatory Exclusion? To answer this question, let's turn to Kim's influential treatment of Explanatory Exclusion.

To begin, we should say that Kim does not provide a systematic characterization of either what should count as a complete explanation or when two explanations should be judged independent. He explicitly trusts that these notions should be satisfactorily elucidated through the consideration of various *prima facie* uncontroversial cases;<sup>21</sup> and a first suggestion that these cases bring repeatedly to light is that a complete explanation must mention a sufficient cause of the explanandum. In this respect, complete explanations would oppose to partial explanations (and, therefore, to partial causes) where 'only a part, perhaps an indispensable part, of a sufficient cause'<sup>22</sup> is mentioned. Yet, some further elucidation is required to apprehend the notion of complete cause which complete explanations presuppose and, in relation, the special role reserved for cases of overdetermination.

The balloon explodes because it is simultaneously hit by two darts. Here we have an effect E and two seemingly independent causes C and C\*, such that E would have still exploded even if either C or C\* had not occurred. In this situation, both C and C\* appear as sufficient causes of E, and they seem to constitute two independent and complete explanations of the same event. Thus construed, we must accept that overdetermination cases are just an exception to Explanatory Exclusion and, as Kim intimates, one might be forced to 'exempt all overdeterminative cases from the requirement of Explanatory Exclusion'.<sup>23</sup> Nevertheless, Kim presumes that this exemption does not constitute a serious problem for Explanatory Exclusion, since most people take for granted that overdetermination does not abound in nature, that cases of overdetermination must be quite rare and exceptional. In fact, Explanatory Exclusion may be conceived as a way of expressing the latter metaphysical assumption.

There is a widespread consensus in current literature that overdetermining causes cannot abound; but there is not so much agreement about how this metaphysical fact should be explained. Many philosophers reject as sheer absurdity that overdetermination could be a general feature of causation.<sup>24</sup> Moreover, some of them have associated, as we shall see, the rarity of overdetermined effects with the very concept of cause, since they envisage the direction of causation as metaphysically linked to the asymmetry in overdetermination, that is, to the fact that causes do not overdetermine their effects while effects do overdetermine their causes. Be it as it may, the rejection of overdetermination, at least the denial that overdetermination could be a general feature of causation, plays a crucial role in the standard physicalist argument about mental causation. In fact, we are happy to accept that causes are individuated so that massive overdetermination is excluded. The problem is

that, if our later arguments are right, the dominant view will be unable not only to justify, but also to respect this metaphysical conviction. In any event, the dominant attitude towards overdetermination imposes a further requirement on complete causes. Complete causes must not only be sufficient for their effects but also (insofar as overdetermination has to be excluded) necessary for them. So the dominant physicalist argument takes for granted that most causal processes (namely, those where overdetermination is avoided) require the existence of causal chains satisfying the following constraints:

- (a) they must be sufficient causes for their effects, that is, they must determine their effects or, to avoid any commitment to determinism, the chances of their effects.<sup>25</sup>
- (b) And, in the circumstances, this sufficient cause must also be necessary for the effect and, consequently, every component of the complete cause must be necessary for the effect. In other words, it must be true that, in the circumstances, the effect would have not been produced if any element of the complete cause had been absent.

Now, we are in a position to examine how causal chains (1) and (2) interrelate. If overdetermination is to be conceived as rather exceptional, then it is clear that cases of mental causation could hardly be viewed as generally overdetermined, and Explanatory Exclusion is of application. It follows then that *c*'s having content property *M* and *c*'s having neural property *N* cannot constitute two independent complete causes of *e*'s having *B*. As a result, we cannot envisage (1) and (2) as two independent explanations. Moreover, Physical Closure emphasizes that (2) must be true, that is, it must be true of some *N* that *c*'s having *N* causally explains *e*'s having *B*.

Hence, we are forced to say that *e*'s having *B* would have occurred even if *c*'s having *M* didn't, unless the latter was dependent on *c*'s having *N*. Hence, if *c*'s having *M* is conceived as independent of *c*'s having *N*, then *c*'s having *M* is bound to be causally inert, since Explanatory Exclusion (together with Physical Closure) precludes the existence of such an independent causal chain. Consequently, if the mental properties of an event are to escape epiphenomenalism, they must depend on its physical properties.

It might be argued against the previous line of reasoning that chains (1) and (2) do not adequately describe those situations where mental causation is actually involved. Thus, some people would point out that,

as we can distinguish in *c* two properties, namely, a neural property *N* and a content property *M*, we can also discriminate in *e* two properties, that is, a behavioural property *B* and an intentional property *I*. So the explananda of both explanations must be different: the neurophysiological state of my brain explains that my arm rises, while the mental property explains that *I* (intentionally) raise my arm.<sup>26</sup> Consequently, we would obtain the two following causal chains:

- Case II:** (1\*) *c*'s having *M* causes *e*'s having *I*  
 (2) *c*'s having *N* causes *e*'s having *B*

Chains (1\*) and (2) certainly have two distinct explananda whereby they remain unaffected by Explanatory Exclusion. Hence, it sounds as if, by splitting the explanandum, we could obtain causal efficacy for the mental properties of an event without dependence upon its physical properties.

The trouble with splitting the explanandum seems to be that this strategy either leads to epiphenomenalism or simply postpones the question. To put it in a nutshell, if epiphenomenalism is to be avoided, Case II is only apparently different from Case I. Either we assume that *c*'s having *M* causally interacts with the physical world, or we assume it does not interact. On the latter horn, we are being committed to a strong version of dualism which rules out mind-body interaction and is, therefore, incapable of accounting for our most manifest intuitions. An initially more sensible version of dualism should concede, on the contrary, that *c*'s having *M* interacts with the physical world. But, in that case, *c*'s having *M* must be the cause of another event *e*\*'s having a certain physical property *B*, which also has a physical cause. Thus, in our example, the mental chain:

- (1\*) *c*'s having *M* causes *e*'s having *I*

requires the development of a causal chain:

- (1\*\*) *c*'s having *M* causes *e*\*'s having *B*.

Hence, to honour Physical Closure, (1\*\*) entails that there must be a neural property *N* such that:

- (2\*\*) *c*'s having *N* causes *e*\*'s having *B*.

But causal chains (1\*\*) and (2\*\*) constitute a Case I. For a causal physicalist, then, every Case II presupposes a Case I. Hence, we can still sustain the thesis that a causal physicalist will only be able to recognize the causal efficacy of 'c's having M' if it depends on 'c's having N'. Consequently, it seems that the only way in which a causal physicalist may recognize the efficacy of mental properties is by accepting the dependence of those properties upon physical properties. This represents a first step in the process of elucidating what, in the light of the dominant stance, the physicalist constraints are for a property to be causally efficacious. Strong Supervenience constitutes the second step.

### 3 Strong Supervenience

The dependence of c's having content property M upon c's having neural property N is, according to the dominant view, the only way to preserve the causal efficacy of the former. This remark must, then, be taken as the first contribution of the dominant view to our initial question about how Will's mental processes interrelate with the corresponding underlying physical mechanisms. It seems that, once the basicness of the physical process is assumed, mental properties will only turn out to be efficacious if they are adequately dependent upon the corresponding physical/neurophysiological properties. We need now to examine how the dominant view scrutinizes these metaphysical images in order to determine how strong such dependence should be. This examination will lead us to formulate Strong Supervenience, as a principle in which, according to the dominant picture, a significant part of our physicalist intuitions coalesce.

Identity is the most extreme form of dependence, but goes beyond what is required. Although an object's having the property 'being a carburettor' depends on its having certain physical properties, 'being a carburettor' is not identical with such a set of physical properties. Since another object, with quite different physical properties, might still possess the property 'being a carburettor'. And, *prima facie*, this situation causes no physicalist strain. What this suggests is that all we should expect is a unidirectional kind of dependence: the tokening of some physical properties being a sufficient (but not necessary) condition for the tokening of 'being a carburettor'. It could, then, be said that the possession by an object of the property 'being a carburettor' *supervenes* upon the possession by that object of a given set of physical properties. Derivatively, this set of physical properties will constitute, in that

object, *the supervenience base* of its possessing the property 'being a carburettor'.

Nevertheless, this is still too vague, since, as defined, the supervenience relation admits of distinct degrees of strength, not all of them equally acceptable for a causal physicalist. In order to determine the strength that, according to the dominant view, actually fits our physicalist intuitions; let us consider, following Kim, three modes of supervenience: Weak, Strong, and Global Supervenience.

*Weak Supervenience* has to do with the idea of mere asymmetric covariation which, as we shall see, fails to grasp the required kind of dependence. So, being 'A' and 'B' names of different families of properties:

'A *weakly supervenes* on B if and only if necessarily for any property F in A, if an object x has F, then there exists a property G in B such that x has G, and if any y has G it has F.'<sup>27</sup>

It is worth stressing that the sort of supervenience relation that Weak Supervenience contemplates is not valid across possible worlds, but only relative to a particular world. As a result, it is obviously compatible with Weak Supervenience that, for instance, in world  $w(1)$ , mental property  $M(1)$  should supervene on physical property P, while, in other worlds  $w(i)$ , an entirely different mental property  $M(i)$  should supervene on the same physical property P. And this might apply to all the physical and mental properties of any given world. This upshot is most dissatisfying, since we do find these possibilities metaphysically perverse. In fact the previous arguments about the dependence between mental and physical explanations seem to require a more robust link.

To illustrate the point, take the functional property of 'being a sparking plug'. We are prepared to explain causally some features of a car's behaviour by the ignition of its sparking plugs. We view 'being a sparking plug' as a causally efficacious property, as picking up a cluster of causal powers. Now the question is: could this cluster of causal powers be possessed by the object in world  $w(1)$ , but not by its physical counterpart in world  $w(2)$ ? If 'being a sparking plug' were just weakly supervenient upon the physical properties of an object, then the answer ought to be affirmative. Yet, according to the dominant view, our metaphysical intuitions appear to impose the opposite response: it could not be so. If an object has the cluster of causal powers that individuate the property 'being a sparking plug', it must have this cluster of causal powers across all physically, or nomologically, possible worlds, that is, across all possible worlds in which the laws of nature hold.<sup>28</sup>

It seems, in any event, that the sort of dependence we are looking for should include asymmetric covariation, not only relative to a possible world, but also across all physically possible worlds. This is the sort of dependence upon the physical that seems to be exemplified by an object's possession of properties such as 'being a sparking plug' or 'being a carburettor', or by biological properties like 'being a chromosome' or 'being an enzyme'. This way we shift from weak to Strong Supervenience whereby we are bound to add a second 'necessarily' (i.e. valid across all possible worlds) to the previous definition. So again, being 'A' and 'B' names for different sets of properties:

'A *strongly supervenes* on B just in case, necessarily, for each x and each property F in A, if x has F, then there is a property G in b such that x has G, and necessarily if any y has G, it has F.'<sup>29</sup>

This definition can also be construed as stating a physicalist condition for the causal relevance of properties, when properties B are the kind of properties that are recognized as 'physical'. It is to this physicalist constraint that '*Strong Supervenience*' will now refer.

Strong Supervenience appears to pick up the sort of dependence that the combination of Physical Closure and Explanatory Exclusion was calling for. Consequently, if the causal efficacy of mental properties is to be consistent with our supposed physicalist intuitions, the mental properties of a system must strongly supervene upon its physical properties. None the less, some people feel reluctant to concede Strong Supervenience because they think that, for the effect of apprehending how the physical and the mental interrelate, a less demanding kind of supervenience will do. To this end, *Global Supervenience* is elaborated, and emerges as the physicalist constraint that the following definition expresses, when 'B' stands for physical properties:

'A *globally supervenes* on B just in case worlds that are indiscernible with respect to B ("B-indiscernible", for short) are also A-indiscernible.'<sup>30</sup>

Undoubtedly, Global Supervenience captures part of our metaphysical intuitions. Two physically indistinguishable worlds ought to share all their mental, moral, aesthetic properties. But this falls short of what is required, since Global Supervenience is consistent with a number of counterintuitive situations.

Firstly, Global Supervenience does not rule out the possibility of two worlds  $w(1)$  and  $w(2)$  differing in the most trifling physical respect, and,

nevertheless, having no mental property in common: one of these worlds including, for instance, consciousness, but not the other.<sup>31</sup> Secondly, in the very same world, two physically identical individuals might possess entirely distinct mental properties.<sup>32</sup> What these possibilities suggest is that Global Supervenience does not succeed in capturing the sort of dependence that Physical Closure, in combination with Explanatory Exclusion, demands. Furthermore, Kim stresses that Global Supervenience without Strong Supervenience becomes a 'brute and unexplainable fact'. How could two physically indistinguishable worlds be indistinguishable in every other aspect if not by the existence of more particular connections, of the sort postulated by Strong Supervenience, between the physical and the rest of the properties of the world? Thus, 'if, as Strong Supervenience affirms, there should exist appropriate connections between specific psychological properties of objects and their physical properties, that would give us a basis for an explanation.'<sup>33</sup>

To sum up, Weak and Global Supervenience apparently fail to grasp the sort of dependence involved in Physical Closure (once Explanatory Exclusion is granted). For they allow for situations that seem to violate certain metaphysical intuitions; only Strong Supervenience appears, on the dominant stance, as an appropriate elucidation of such intuitions. To get a clearer picture of the dominant view, we need to explore some further physicalist principles, which are closely associated with the previous considerations. This is what we purport to do in the next section, while in the final section we plan to consider why mental properties may be at pains to satisfy these physicalist demands.

#### **4 Narrowness and Minimality**

The dominant view tends to see in Strong Supervenience a fairly adequate expression of the physicalist intuitions which derive from Physical Closure and Explanatory Exclusion. These reflections rely, though, on a certain understanding of what a physical property is. Section 4.1 elucidates this notion, where we bring out the ties that Strong Supervenience bears with two other physicalist constraints, namely: Intrinsicness and Narrowness. This will improve our understanding of the physicalist requirements that the dominant view commonly emphasizes. Yet, in section 4.2, we will argue that Physical Closure and Explanatory Exclusion entail a more demanding physicalist constraint, which, despite being crucial, goes generally unnoticed: 'Minimality' is the word that we will reserve for it. In fact, we will close by claiming that the causal physical-



ist is bound to recognize that only those nonbasic properties that satisfy both Narrowness and Minimality may count as causally efficacious.

#### 4.1 *Physical properties: Intrinsicness and Narrowness*

At first sight, one might claim that history and context are causally irrelevant unless they have left some trace in the present state of affairs. We often explain a teenager's aggressive behaviour by the hard circumstances in which she has been raised. We can report those circumstances including numerous details, indicating particular persons, actions, and places. But all these historical considerations would be utterly irrelevant from the viewpoint of explaining the teenager's behaviour, if we did not assume that they have transformed her personality in a certain way, and left a lasting, specific mark in her brain. Environments of different kinds may foster a similar aggressive behaviour in children, while distinct children may react differently in the same kind of environment. It seems, then, that what accounts for the current hostile or friendly behaviour of a teenager are not the hardships or benevolence of the environment where she was actually brought up, but the way her personality has been altered by that environment and, in terms of the causal physicalist, the kind of neurological structure that repeated exposure to a certain environment has wired in her brain.

What are, however, the intuitions about physical properties and causation that render these remarks plausible? Let us begin by considering our intuitions about the behaviour of two physical duplicates A and B. It seems clear that, for a spatiotemporal context C which fixes the possession of a property P by A, it is true that had B been placed in the same context as A, then B would also have had property P. This intuition is so undemanding that it holds even when property P is a relational property, e.g. 'having had a headache yesterday'. For it is true that, if B had been in A's place, in the context C which fixes the possession by A of the property 'having had a headache yesterday', then B would also have had this property, since context C would only fix the possession by A of the property at stake if it should include what occurred to A yesterday.

The previous considerations suggest a sort of *Substitutivity Test* for physical twins:

For any property P and any pair of physical twins x and y, it is true that, when x is in a physically possible context that fixes that x is P, had y been in that context instead of x, y would also have had P.

When people stress the basic character of physical properties, they may be simply claiming that physical twins (i.e. twins regarding physical properties) do pass this Substitutivity Test. In fact, only if physical properties are defined in terms of this test, several physicalist intuitions proclaimed by the dominant view will come into focus. Accordingly, we can say that the idea that physical properties are basic presupposes that:

A *basic* property is a property that belongs to the minimal subset B of the properties of the world that satisfies the following requirement: every two objects that share all their B properties pass the Substitutivity Test.

The idea that physical properties are basic in this sense seems to be equivalent to the principle of Physical Closure. First, it seems true that if we should dismiss Physical Closure, we ought to abandon the idea that physical properties are basic. For the denial of Physical Closure entails (i) that there might be some physical effect that could not be completely explained by its physical antecedent. And this appears to entail (ii) that, sometimes at least, when a physical property is instantiated, the whole physical context is not able to fix the effect or, at least, the objective chances of it. Consequently, it sounds that the totality of the instantiated physical properties is unable to fix all causally relevant properties. And this counts, by definition, as a straight refusal of the basic character of physical properties.<sup>34</sup>

Conversely, Physical Closure seems to entail the basic character of physical properties. If we should repudiate this basic character, we would have to drop Physical Closure with regard to those properties that are able to affect the physical world. If physical properties were not basic, then it might be possible that certain causal powers were not fixed by the instantiated physical properties and their physical contexts. As a result, certain differential effects could not be accounted for in terms of physical differences. But those effects may be physical or nonphysical. If they were physical, such a situation would amount to a straightforward rejection of Physical Closure. If, on the contrary, the effects in question were not physical, they must, *ex hypothesi*, be able to have physical effects. Then the previous argument applies.

At the beginning of this chapter, we made an initial concession to the causal physicalist, namely: there is an intelligible notion of physical property. And, coherently, we renounced asking for a clear specification of what those properties could be. Now we are prepared to make a related

concession. We are prepared to grant that the notion of basic property just introduced in terms of the Substitutivity Test can identify the relevant class of 'physical' properties. Some could, indeed, object that by granting this we are being by far too generous, since we are conceding, for instance, that there is an interesting subset of properties that satisfies the definition of 'basic property', and that the components of such a subset are fixed by the truth of an ideal basic science of nature, namely, physics. In other words, we are allowing the causal physicalist to assume that a stipulation about basic physical properties bears some relevant links with the methodological principles and the empirical findings of physics, and this is certainly a rather liberal concession.

In fact, the definition of a basic property can only be of some help to causal physicalism if it generates a small and interesting enough subset that could fit certain previous intuitions. Given that nobody even attempts to provide a list of properties that would possess the metaphysical virtues that the dominant view ascribes to physical properties, it sounds fair enough to stipulate simply that physical properties are those that could satisfy the relevant metaphysical requirements (that is, being basic in the sense just explained). This stipulation does not exclude by itself that every predicate would refer to a different basic property and, thereby, the risk of emptiness is manifest. If causal physicalists are to avoid this risk, they must be relying on a further assumption, namely: that the set of basic properties (i.e. the set of properties that meet the stipulation at stake) has to be interesting enough or, in other words, has to be fixed by a very small subset of our predicates, say, the predicates used in basic science explanations.

So, the point has to be that there must be a small subset of properties that satisfy the aforementioned stipulation and that might be fixed in terms of something like the causally relevant properties of an ideal basic science. The problem is that scientific explanation seems to be necessarily concerned with the idea of attributing identity of causal powers to different states of systems. It is not by accident, for instance, that physics books accept as explanatory, as causally relevant, properties which allow for several different realizations in terms of other, more fundamental properties. Consider, for instance, the following paradigmatic example of physical explanation: the correlation between variations in the pressure in an ideal gas and variations in its temperature. The law that fixes this correlation (Boyle's law) does not explain the individual behaviour of any given particle, but the overall effect on the pressure variable that is produced by altering the temperature variable.<sup>35</sup> If some would say that, as a result, Boyle's law cannot be a basic law of nature,

then we would be at trouble to understand what could be the form of the basic science of nature.<sup>36</sup>

A different, more sensible possibility would be to construe the notion of 'basic property' as relative. Thus, the set of, say, neurophysiological properties could be basic with regard to the set of, say, mental properties; even if there could be a third set of properties that would count as basic with respect to neurophysiology. Thus, at any stage, the scientific enterprise could then appeal to the causal efficacy of states that are multiply realizable in terms of other properties. Moreover, on this relative notion of basic property, all the previous physicalist arguments for the primacy of neurophysiological states in mental causation could still retain their force. We would not need, then, the idea of a basic science of nature that only appeals to the causal efficacy of the most basic properties of the world.

In any event, our challenge to causal physicalism will not focus on this kind of worry, since, for the sake of argument, we are ready to grant that the causal physicalist would eventually yield satisfactory examples of basic properties. When we shall consider some putative examples of dependence of functional and mental properties upon more basic properties, we will present as relevantly basic the kinds of properties that the dominant view accepts as such: force, mass, shape, electrical patterns, neurophysiological patterns, and so on. We will analyse those cases without discussing the connection that those relatively basic properties bear to the putatively ultimate basic properties of the world. It is clear, however, that, by calling into question the physicalist assumptions about the link between, say, mental and neural properties, we are ipso facto attacking a physicalist understanding of the link between any other kind of property and the putatively ultimate basic properties of the world.

Be that as it may, it is important to highlight that the dominant view tends to consider that a relevant metaphysical principle follows from the previous stipulation of physical properties as basic, namely:

**Intrinsicness:** Two objects that share all their intrinsic physical properties must also have in common all their causal powers across all nomologically possible contexts.

This principle could also be stated in terms of Strong Supervenience. According to Intrinsicness, any causal power of an object is necessarily preserved by any of its physical duplicates. So, there cannot be any difference in causal powers without a difference in the (intrinsic) physical properties of the objects at stake. To put it another way, suppose that

object A is in a context C which fixes the possession by this object of a particular property P. We may say that, in this context, object A has property P. Now consider a physical duplicate B. Since by definition physical properties pass the Substitutivity Test, it follows that object B has the ability to acquire property P if placed in context C. Of course, this ability is something that A and B share, independently of their actual history or location. What the dominant view judges is that those abilities, shared across all physically possible contexts, causally explain the acquisition, in a particular context, of some conditional causal powers. It seems then that those abilities fixed as basic by the Substitutivity Test are themselves causal powers that an object would preserve across all physically possible contexts and, therefore, Intrinsicness ought to be acknowledged.

This transition is usually taken for granted by causal physicalists; yet, we intend to argue that it does not hold. For the transition in question relies on some apparently elementary intuitions about causality and explanation which, if we are right, will turn out to be quite inappropriate. It is commonplace, within the dominant view, to mention Stephen Stich's remarks to the purpose that the principle of substitutivity directly entails the intrinsicness of causal powers.<sup>37</sup> At first sight, this move can hardly be objected, insofar as it uses a notion of 'power' that can be legitimately derived from the everyday meaning of the word. It is obvious that, insofar as a pair of physical twins would respect the Substitutivity Test, they have the following 'power': each of them would act as the other actually does if it were placed in the context in which the other is actually placed. However, it does not follow, and this will be our main objection, that this 'power' is a causal power or, in other words, that the counterfactual that expresses that 'power' should really count as a causal counterfactual.

It is important to notice that we are not asking for a reductive account of the distinction between intrinsic and extrinsic properties. Of course, this is something that the dominant view has not provided.<sup>38</sup> And it is not by accident that no philosopher has succeeded in such an enterprise. Yet, some previous intuitive conception of that distinction is required if Intrinsicness is to be at all an intelligible principle. In everyday life, we do seem to have little problem in drawing the required distinction. The shape of a key is an intrinsic property of it, while its being placed on my table does not count as an intrinsic property of the key. There is a number of alternative ways of expressing this intuition. We could thus say that intrinsic properties depend on the constitution of the thing itself, independently of the properties of any other object. But we could also insist that intrinsic properties are preserved by perfect duplicates of the

thing at stake or, alternatively, that a change in the intrinsic properties is a real change in the object, since a change in the shape is a change in the key, but its change of position on the table does not represent a change in the key. The fact that we change the position of the key does not change the key itself. However powerful these remarks may sound, it is quite clear that they rely on a circle of closely interconnected notions, and there is little hope that any of them would help to provide a reductive account of the rest. Perfect duplicates are just objects that share all their intrinsic properties. To say that certain facts about an object are independent of any other thing amounts to claiming that those facts are fixed by the intrinsic properties of the object in question. A real change in an object is simply a change that does not require any variation in the facts about things that are independent of the object. By rehearsing this circularity, we are not at all suggesting that there is something wrong in these everyday distinctions; still it may serve as a warning against certain metaphysical moves that (like those of the causal physicalist) take for granted that such everyday distinctions can be safely projected onto much more complex situations, without a proper assessment of the conditions under which these distinctions become intelligible in our everyday life.

Be it as it may, Intrinsicness seems to impose a certain interpretation of what a non-intrinsic causally efficacious property has to be. Take, for example, the property of 'weighing half a kilogram', which is a relational property of a book Q: a physical twin of book Q in a different gravitational context would have another weight. What seems indisputable, though, is that this extrinsic property is determined by the basic properties of the book and the basic properties of the context in which book Q is placed. A physical duplicate of Q and a physical duplicate of the gravitational context in which book Q is placed would fix the very same weight for the book. Furthermore, it seems clear that the book's property 'having a weight of half a kilogram', which is relational regarding book Q to which we apply the predicate 'heavy', may be intrinsic with respect to a larger system that should encompass both book Q and the corresponding gravitational field. For a relational property R of Q must strongly supervene upon the basic properties of a broader system. From this perspective, the distinction between extrinsic or relational and intrinsic or internal properties can hardly be construed as a distinction between types of properties, since the cluster of causal powers that any property picks up may appear as either extrinsic or intrinsic, depending on how the bearer of such causal powers is individuated.<sup>39</sup>

In any case, what Intrinsicness demands is that there must be a specific way in which Q contributes to the tokening of R within a larger system. The ability to contribute in a certain way can be envisaged as an intrinsic property of an object Q, and what fixes this contribution is the narrow correlate of the relational property R in object Q, which can then be defined as:

the specific way in which the basic properties of object Q contribute to the tokening of relational property R in the relevant context.

So, to follow up our example, we could say that an object's mass is an intrinsic property of it, a narrow correlate of its having certain weights when placed in the suitable gravitational fields. And what Intrinsicness implies is that:

**Narrowness:** Each nonintrinsic property of an object must have its narrow correlate.<sup>40</sup>

In fact, Narrowness and Intrinsicness are, on the dominant view, equivalent, since not only does Intrinsicness entail Narrowness, but also the converse is surely true as well. For, if extrinsic properties have narrow correlates, then physical counterparts will certainly have all their causal powers in common, and this is what Intrinsicness amounts to.

If Narrowness is to be granted, then science and genuine explanations should not violate the kind of taxonomy imposed by the maximally counterfactual notion of causal power that, as we saw, generates Intrinsicness. The relational properties of an object pick up causal powers only if they have narrow correlates.<sup>41</sup> In Chapter 3, we purport (a) to demonstrate that there are no such things as narrow correlates; and (b) to challenge the notion of causal power which permits the dominant view to derive Intrinsicness from the definition of physical or basic property in terms of the Substitutivity Test.

## 4.2 *Minimality*

The dominant view maintains that the combination of Physical Closure and Explanatory Exclusion imposes that event c's having M should depend on c's having N, and the exploration so far presents Strong Supervenience as expressing the sort of dependence involved. In this

section, we intend to argue, however, that Explanatory Exclusion and Physical Closure actually call for a more stringent constraint, which we shall call 'Minimality'.

To begin we must recall that Explanatory Exclusion is committed to the assumption that cases of overdetermination are rare. And, trivially, insofar as massive overdetermination is to be averted, one cannot in general accept that two different complete causes are responsible for the same effect. Yet that is exactly what would happen, if there could usually be two alternative bases of supervenience that were causally efficacious when a particular instantiation of a nonbasic property is causally active. It seems then that the only way to avoid overdetermination when a nonbasic property *S* is causally efficacious is by requiring the existence of a *single* set *P* (1 . . . *n*) of physical properties that is a minimal supervenience base, that is, a supervenience base that comprises no superfluous or idle element. More strictly, a *minimal supervenience base* could be defined as follows:

*P*(1 . . . *n*) is a minimal supervenience base of a property *S* if and only if *P*(1 . . . *n*) is a supervenience base *S*, and each *P*(*i*) in *P*(1 . . . *n*) is a necessary condition for the latter to be such a supervenience base.<sup>42</sup>

To give an example, suppose we have a particular orange squeezer with a panoply of physical determinations. Of course, its functional property 'being an orange squeezer' does not depend on the totality of its physical determinations. The colour and even the exact number of teeth on the surface that gets in contact with oranges is irrelevant. A minimal supervenience base would disregard these superfluous elements, and only include those physical determinations of this particular orange squeezer in virtue of which it has the functional property 'being an orange squeezer', so that each of these physical determinations should be a necessary component of the supervenience base which they jointly represent.

Now, it is easy to see that only this requirement ('*Minimality*', hereafter) of a single minimal base of supervenience for each instantiation of a nonbasic property deters the spread of overdetermination, and respects the fundamental intuitions of the causal physicalist as to the causal relevance of nonbasic properties. Suppose that we have a set *P*(1 . . . *n*) of physical properties that is a nonminimal supervenience base of a particular instance of *R*. Hence, there must be at least one physical property *P*(*i*) without which the previous set would still be a supervenience base of this particular instance of *R*; so that we



can dispense with it and still R would be instantiated. We would have then two different supervenience bases instantiated on this particular occasion:

- (1)  $P(1 \dots n)$
- (2)  $P(1 \dots n) - P(i)$

If overdetermination should not proliferate, we cannot consider that these two supervenience bases are causally responsible in the same sense. It sounds, then, natural to say that, if (2) is enough for the actual instantiation of R, then the causal efficacy of (1) – at least, as an instance of R – depends on its having (2) as an ingredient of it. So, we are bound to recognize that, in this situation, it is supervenience base (2) that actually does the causal job.

But, let us motivate Minimality in a more formal way, out of the fundamental principles of causal physicalism, namely: Physical Closure and Explanatory Exclusion. This line of reasoning will demonstrate how the demand of single minimal bases of supervenience is connected to the notion of a complete cause involved in those principles. We have seen that, on Physical Closure, e's having physical property B must have a complete physical explanation. Specifically, there must be, in this view, a cause that is complete in such a way that each of its components is necessary for it to count as a sufficient condition of the effect. By contrast, we are faced with a partial cause when only a necessary part of a sufficient condition is mentioned. In any case, Explanatory Exclusion presupposes that:

A necessary condition for c's having N being a (partial) cause of e's having B is that the former is a necessary part of a sufficient condition for the latter.<sup>43</sup>

Let us now suppose, as in our example in section 2, that 'M' names a mental property causing an effect B, that 'N' stands for a neurophysiological property that also causes B, and that 'B' designates a certain behavioural property. According to Explanatory Exclusion, c's having N would not count as a cause of e's having B if the former were a superfluous component of a sufficient condition for the occurrence of the latter; and similarly for c's having M. In that case, it is easy to see that for N to be causally relevant, it must be a minimal cause of B, i.e. a cause that is deprived of superfluous ingredients. Imagine that N is a non-minimal cause of B. Thereby e would still have B even if, instead of

N, an alternative property  $N^*$  were tokened. Just individuate  $N^*$  as N minus N's superfluous ingredients. But, of course,  $N^*$  is in fact tokened, and by itself guarantees, in the circumstances, the tokening of B. As a result, c's having N cannot be a cause of e's having B because it includes elements which are superfluous regarding a sufficient condition for the bringing about of the effect. The causal role of N is consequently overridden by  $N^*$ .

All this comes to emphasize that Minimality is an overall physicalist demand that can be equivalently expressed both in terms of supervenience and in terms of causation. In the latter terms, Minimality would demand that whenever a property P is causally responsible of an effect E, there must be a single minimal set of basic properties that is a complete cause of E. In terms of supervenience, the requirement imposes that every instance of a nonbasic property P should have a single minimal supervenience base with no superfluous elements in it: a supervenience base all of whose components are necessary for the instantiation, in the circumstances, of the supervenient property.

All this surely reveals something about the way the particular instances of N and M are related. As we have seen in section 3, Physical Closure and Explanatory Exclusion straightforwardly entail that whatever the causal efficacy of M may be on a particular occasion, it cannot go beyond the causal powers of the physical causal chain that, on this particular occasion, includes N as a necessary ingredient and is enough to causally explain B. What our recent remarks highlight is that, if N has to contribute to fix the causal efficacy of M on a particular occasion, then N has to be a necessary ingredient of a single minimal base of supervenience of M and, therefore, a nonsuperfluous member of the minimal set of basic properties that, on this particular occasion, is the complete cause of B.<sup>44</sup>

This upshot is quite neutral with regard to, for instance, the controversial issue among functionalists as to whether mental (functional) causation is just neurophysiological (physical) causation or not. So, our argument for Minimality does not take sides in this dispute. Independently of the current debate about how the causal powers of supervenient properties are exactly linked to the causal powers of subvenient properties, there is a widespread agreement within the dominant view that the two following principles hold: (i) whenever a particular instance of a mental property is causally efficacious there is a set of properties (neurophysiological, physical) that constitute the basis of supervenience of that mental property on that particular occasion; and (ii) the causal efficacy of the mental property on that given occasion is fixed by the causal efficacy of this particular supervenience base. Even

conventional versions of functionalism seem unintelligible without this minimal assumption. A conventional functionalist, even if she denies that supervenience is the whole story about how our mind is able to causally affect the world, is pleased to recognize that the supervenience of mental properties upon physical properties is a *sine qua non* requirement for mental causation. Supervenience is, in Fodor's words, 'the best idea anyone has ever had about how mental causation is possible'.<sup>45</sup>

Be that as it may, our challenge to Minimality will turn out to be crucial in this book. We intend to argue, in Chapter 3, that Minimality is after all self-defeating as an attempt to characterize the causal efficacy of functional properties, since these properties are unable to fulfil that demand. Furthermore, in Chapter 5, we will insist that causal physicalism relies on some views about causation which involve Minimality, namely, which are only intelligible as expressions of the demand of minimally complete causal explanations. Yet, the conviction that such a demand can be satisfied in terms of physical or neurophysiological properties will be dismissed as completely implausible. In fact, we shall argue that sticking to Minimality would force us to recognize overdetermination as an extended phenomenon. But, as we have seen, the entire physicalist argument on mental causation crucially depends on the refusal of massive overdetermination.<sup>46</sup>

## 5 Narrowness and Mental Causation

It is an important part of our intuitions about the causal efficacy of the mental that mental states have got their causal powers in virtue of their respective content. This is an intuition that both the dominant view and us would like to honour. Yet, according to causal physicalism, mental contents can only be causally relevant if they would satisfy certain physicalist constraints, namely: Narrowness and Intrinsicness. The problem arises, however, that there seems to be powerful arguments to the effect that, at least our everyday mental contents, are individuated relationally. It is not an accident, then, that a lot of work in the philosophy of mind had been devoted, during the last decades, to seek to reconcile the causal relevance of mental contents with the preceding physicalist considerations in favour of the intrinsic character of causal powers. So, let us briefly sketch the standard arguments for the relational character of ordinary mental contents, and indicate the way in which such arguments have been traditionally approached within the dominant view.

We may begin by rehearsing the thought-experiments proposed respectively by H. Putnam and T. Burge.<sup>47</sup> Putnam's piece of science fiction includes a hypothetical Twin-Earth, where a physical counterpart of Oscar dwells. Twin-Earth is identical to Earth in all respects except in those that derive from the fact that a liquid that has the same macroscopic properties as water on Earth possesses a different molecular structure, say, XYZ. The question is whether Oscar's beliefs and desires concerning water and his twin's beliefs and desires concerning Twin-water ('twater', for short) have the same content. Many philosophers have envisaged Putnam's reflections as inducing a negative answer to this question, even though Putnam, in 'The Meaning of Meaning', does not explicitly conclude that those mental contents that explain Oscar's and Twin-Oscar's behaviour should be externally constituted. For he just focuses on the meaning of natural kind terms such as 'water', 'gold', 'silver', and so on, and argues that natural kind terms include a hidden indexical component: water is the stuff that bears a certain similarity relation to water around here. Water at another time or in another place or even in another possible world, has to 'bear the relation same-L to our "water" in order to be water'.<sup>48</sup> It follows then that, in Twin English, 'water' does not mean water. On the face of this, one may feel tempted to explore how these semantic considerations may affect the content of the mental states expressed on Twin Earth by the term 'water'. And some may be encouraged to conclude that people's thoughts on Twin Earth do not refer to water. For, otherwise, one would have to defend an implausibly sharp distinction between the content of a certain kind of belief and the meanings involved in their most natural linguistic expressions. An externalist would typically urge that it is precisely because water-thoughts and twater-thoughts are concerned with different substances, with disparate extensions, that the contents of such thoughts differ too.

Burge's thought-experiment is meant to expand the scope of this kind of externalist argument. For Putnam's cases only seem to affect those thoughts that somewhat refer to natural-kinds, thereby only some of our thoughts would have been proved to be extrinsically individuated. By contrast, Burge's examples extend this upshot to all common concepts by taking into consideration variations not only in the natural, but in the social environment as well. Suppose that Will has a certain notion of what arthritis is. He knows that arthritis may be painful, but that it is better to have arthritis than cancer, and so on. Let us assume that he has plenty of true beliefs about arthritis, even if he does not know every single truth about arthritis. In such a case, Burge argues, we are inclined to think that Will possesses the concept of arthritis and that he is in a

position to have arthritis-thoughts. Something must be added to the story, however: Will falsely believes that he has developed arthritis in his right thigh. This belief is false because, as a matter of social convention, 'arthritis' only applies to a painful inflammation at the joints.

At this stage, Burge proposes a counterfactual situation. Twin-Will has exactly the same story as Will, and their corresponding social environment coincides in every respect but one: in Twin-Will's environment, 'arthritis' also applies, as a matter of social convention, to the inflammation of thighs. In such a case, Twin-Will truly believes that he has developed twin-arthritis in his thigh. Arthritis-thoughts and Twin-arthritis-thoughts will therefore differ in content. We accept in the initial description that, despite Will's false belief about having arthritis in his right thigh, Will had a background of abilities sophisticated enough to say of him that he has arthritis-thoughts. And this can be so because commonsense ascriptions of contents allow for an incomplete understanding of them. By similar reasons, Twin-Will does have Twin-arthritis-thoughts, but not arthritis-thoughts. It becomes apparent, then, that, despite the fact that Twin-Will and Will coincide in all their intrinsic properties, their thoughts have different contents because they are placed in disparate social environments. Mental contents are individuated relationally, that is, they may vary in accordance with environmental changes, even if the intrinsic properties of the organism that possesses them remains unaltered.

The dominant view has traditionally argued that these arguments only affect our commonsense theory of content, and that they do not rule out the possibility of developing a technical notion of content that were more suitable from the viewpoint of causal explanation. If workaday contents are broad, relational; a notion of narrow, intrinsic content ought to be elaborated.<sup>49</sup> As a result, a variety of notions of narrow content have been proposed throughout the years.<sup>50</sup> In fact, this strategy seems to fit the requirements imposed by Narrowness, according to which a relational property of an object *O* only picks up causal powers if it has a narrow correlate, which fixes the object's contribution to the tokening of the relational property.

Chapters 2 and 3 argue, by contrast, that not only mental properties, but also functional (and, in general, nonbasic) properties are at pains to comfort to Narrowness and, relatedly, to Minimality. If we would finally succeed, then we would be in a position to benefit from a rather general argument against all putative theories of narrow content, as well as against any functionalist account of the causal efficacy of nonbasic properties. This being so, we will, at a later stage, proceed to revise the metaphysical motivation for Narrowness and Minimality, and end by

proposing an alternative analysis of how basic and nonbasic properties interrelate, which, as we shall see, poses no problem to the causal responsibility of either functional or mental properties. Let us, then, begin our challenge to Narrowness.

# Minds, Causes, and Mechanisms

*A Case Against Physicalism*

Josep E. Corbí and Josep L. Prades

 **BLACKWELL**  
Publishers

Copyright © Josep E. Corbí and Josep L. Prades 2000

The right of Josep E. Corbí and Josep L. Prades to be identified as authors of this work has been asserted in accordance with the Copyright, Design and Patents Act 1988.

First published 2000

2 4 6 8 10 9 7 5 3 1

Blackwell Publishers Ltd  
108 Cowley Road  
Oxford OX4 1JF  
UK

Blackwell Publishers Inc.  
350 Main Street  
Malden, Massachusetts 02148  
USA

All rights reserved. Except for the quotation of short passages for the purposes of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying or otherwise, without the prior permission of the publisher.

Except in the United States of America, this book is sold subject to the condition that it shall not, by way of trade or otherwise, be lent, resold, hired out, or otherwise circulated without the publisher's prior consent in any form of binding or cover other than that in which it is published and without a similar condition including this condition being imposed on the subsequent purchaser.

*British Library Cataloguing in Publication Data*

A CIP catalogue record for this book is available from the British Library.

*Library of Congress Cataloging-in-Publication Data*

Corbí, Josep E.

Minds, causes, and mechanisms : a case against physicalism / Josep E. Corbí and Josep L. Prades.

p. cm. — (Aristotelian Society monographs series; 17)

Includes bibliographical references and index.

ISBN 0-631-21801-7 (alk. paper). — ISBN 0-631-21802-5 (pbk. : alk. paper)

1. Causation. 2. Philosophy of mind. I. Prades, Josep L., 1954- .

II. Title. III. Series.

BD541.C67 1999

122 — dc21

99-16131

CIP

Typeset in 10 on 12 pt Plantin

by Best-set Typesetter Ltd., Hong Kong

Printed in Great Britain by MPG Books Ltd, Bodmin, Cornwall.

This book is printed on acid-free paper.



# Contents

Acknowledgements	ix
Introduction	xi
<b>1 Physicalism and the Mental: The Dominant View</b>	<b>1</b>
1. The Dominant View in the Debate on Mental Causation	4
2. The Causal Closure of the Physical World and Explanatory Exclusion	16
3. Strong Supervenience	21
4. Narrowness and Minimality	24
5. Narrowness and Mental Causation	35
<b>2 An Initial Tension: Narrowness and Multiple Realization</b>	<b>39</b>
1. Multiple Realization	40
2. A Critique of Fodor's Notion of Narrow Content	45
3. Functional Properties and Narrowness: Some Initial Remarks	51
<b>3 Dispositions, Minimality, and Intrinsic Causal Powers</b>	<b>58</b>
1. Twin-Substitutivity and Narrow Correlates	59
2. Multiple Realization and Minimality	61
3. Realizing Contexts and Sets of Circumstances	67
4. Dispositions and Meta-Dispositions	73
5. Basic Causal Functions: Conditional vs. Unconditional Dispositions	78
6. Normality Conditions and Causal Dispositions	88
7. Dispositions and Intrinsic Bases	92
8. Supervenience Constraints	100

<b>4</b>	<b><i>Ceteris Paribus</i> Laws and the Autonomy of Nonbasic Properties</b>	109
1.	The Dominant View: A First Approach to <i>Ceteris Paribus</i> Laws	111
2.	A Preliminary Doubt	114
3.	The Fork Problem	120
4.	An Alternative Approach to <i>Ceteris Paribus</i> Laws	128
<b>5</b>	<b>Strict Laws, Causes, and Background Conditions</b>	135
1.	Causal Relata: Davidson's Theory of Causation	137
2.	Complete Causes and Causal Explanation	144
3.	Strict Laws and Causal Counterfactuals	151
4.	Overdetermination and the Individuation of Causes	159
5.	An Overall Antireductionist Argument	167
6.	Background Conditions, Normality, and Interference	178
7.	The Objectivity of Causal Facts and the Absolute Conception of Reality	185
<b>6</b>	<b>Mental Causation</b>	193
1.	Physical Closure, Overdetermination, and Mental Causation	195
2.	Implementing Mechanisms and Mental Causation	202
3.	Tracking the World	212
4.	The Externalism of the Mental	227
	Notes	233
	References	252
	Index	260