

Introduction

On 13 April 1865, John Wilkes Booth shot Abraham Lincoln. It happened in a theatre near the White House, at about 10.13 p.m. Booth wanted to kill Lincoln in retaliation for his treatment of the South during the civil war. So upon learning that the President would attend a play that night, Booth (an actor who knew the layout of the theatre) decided to hide near the President's box and wait for an opportunity to get inside. At some point, the Secret Service guard left his station to get a better view of the stage. Booth saw his chance, got into position, and fired his pistol.

American history texts usually contain some such account of the events that led to Lincoln's death. Taken at face value, the brief account just given offers a description of Booth's action, and an intentional explanation of that action. This explanation reveals why Booth acted as he did, by providing information that makes his reason for action apparent, even if we find the reason insufficient (and the action deplorable). Following Davidson (1963, 1967*a*, 1971) and many others, I hold that rationalizing explanations cite causes, and that actions are events. An intentional explanation of a person's action cites a mental cause of the event that is his action. When we speak of mental *causes*, we may be speaking of various things: events, like Booth's coming to see that he had an opportunity to kill Lincoln; states, like Booth's desire to kill Lincoln; or perhaps other entities, like the fact that Booth wanted to kill Lincoln. But I will focus primarily on events, which have the effects they do, in part because of prevailing background conditions.

Given that actions have mental causes, a variety of considerations make two further claims very tempting: paradigmatic actions are bodily motions; and the mental causes of human actions are certain biochemical events in our brains. On this view, Booth's action of pulling the trigger was a motion of his trigger finger; and when we say 'Booth saw his chance', we are citing a biochemical cause of that finger motion. But I think these further claims are false. My principal aim is to defend an alternative conception of intentional explanation, according to which actions typically cause bodily motions, and the mental causes of our actions are not biochemical events of any sort. (Let me say at the outset that I have been influenced by Hornsby (1980, 1997), in elaborating and modifying certain aspects of Davidson's view.)

1

My plan is to argue first that events like Booth's action of pulling of the trigger cause events like the motion of Booth's trigger finger. With this conclusion in place, I develop a conception of intentional explanation that leaves it open whether (human) mental events are really biochemical events, or whether a non-Cartesian form of dualism is true. Finally, I contend that the latter option is better, all things considered. Along the way, the following theses will be defended: among a person's mental events are his *tryings*—events that typically cause bodily motions, and are caused by (the acquisition of) beliefs and desires; paradigmatic actions are tryings; an intentional explanation usually cites a belief or desire that causes an action/trying that causes a bodily motion; and human mental events are causes distinct from, although they supervene on, biochemical events. In this sense, both reasons and actions belong to a distinctively mental domain. To explain a person's action is (at least in the first instance) to explain a mental episode of trying. Some of these claims may initially sound paradoxical. So it may be helpful to provide a less telegraphic overview.

Chapter 1 begins with a puzzle that arises in giving a semantics for action sentences. I think Davidson was basically right about the logical forms of sentences like 'Booth shot Lincoln' and 'Booth pulled the trigger'; they involve quantification over events, like shootings and pullings. Given this analysis and some further facts, it follows (perhaps surprisingly) that the event of Booth's shooting Lincoln was distinct from the event of Booth's pulling the trigger. But as Davidson noted, there are powerful reasons for saying that 'the shooting' and 'the pulling' can be used to describe a single action. I argue that the best resolution of this puzzle, which has implications for several issues in semantics and action theory, treats event descriptions like 'his shooting of Lincoln' as ambiguous: they can be used in referring to actions/tryings, or to complex events (processes) that begin with an action and end with some effect of the action (cf. Thomson 1977). The shooting and the pulling are distinct complex events; but there was one action, which caused Booth's finger motion, which caused the motion of the bullet.

It may seem odd to begin a study of intentional explanation with a puzzle about action sentences. But this provides a good forum for expediting and testing the coherence of my views about actions. Moreover, in discussing intentional explanations of why *Booth pulled the trigger*, one would like to know how the italicized sentence is related to Booth's action. The puzzle reminds us that this relation is not obvious. Talk of events also raises the question of what events are; but I do not want to begin with a criterion of individuation for events. Such criteria are tendentious, and I

want to explore (not prejudge) issues about which event descriptions describe the same events. Davidson's event analysis, however, can be defended on independent grounds. The conception of events implicated in ordinary speech is quite rich. So we can usefully begin by saying that events are what sentences that quantify over events quantify over. I say a little more in subsequent chapters. But the idea, familiar in the analytic tradition, is to use semantics as a prolegomenon to metaphysics.¹

That said, the meanings of action sentences do not settle the questions of interest here. While the semantics is highly suggestive, further argument is needed to show that actions cause bodily motions. And a full defence of the claim that actions are tryings requires an account of tryings. Eventually, I offer a functional account (similar to O'Shaughnessy (1973, 1980)), in terms of how tryings are nomically related to beliefs and desires. In the interim, I operate with a pretheoretic notion of trying. In particular, I assume that every trying has a *content*—a propositional specification of its success conditions.

In this respect, a trying is like the acquisition of a belief (which can be true or false) or a desire (which can be satisfied or not). For familiar reasons, contents appear to be more finely grained than states of the mind-independent world. Suppose Booth once met a nice man called 'Abe', who later became president. Then intuitively, Booth could have believed that Lincoln was in the theatre, without believing that Abe was in the theatre, even though Lincoln *was* Abe; and Booth could have tried to shoot Lincoln without trying to shoot Abe. This leads to the cluster of puzzles associated with the semantics of propositional attitude ascriptions. I approach the question of what contents are via these puzzles.

In Chapter 2, I defend a version of Frege's view: epistemic verbs like 'believe' ('want' and 'try') express relations between thinkers and *senses* of sentences; 'that P' is a device for referring to the sense of 'P'; and verb phrases like 'try to shoot Abe' contain a clause relevantly like 'that he shot Abe'. But unlike Frege, I hold that terms embedded in 'that'-clauses have their customary referents. Even in the sentence 'Booth believed that

¹ I do not deny that many events are changes (see Lombard 1986). We typically, and perhaps always, see instances of causation as cases in which one change brings about another. And it is tempting to say that events are changes of state in an object. (As I typed these words, my fingers moved, keys went down, and text appeared on my screen; as my fingers changed states, this caused changes of state in the machine.) But some events, like vigils, may not be changes; and some occurrences may not be changes in any object (see Steward 1997; Strawson 1959). In Ch. 3, I address the relation of events to states more explicitly. Some philosophers will find talk of events suspect, absent a substantive account of what it is for there to be *one as opposed to two* of them (see Quine 1969; cf. Brody 1980). Perhaps as Davidson (1969) suggests, $E = F$ iff E and F have the same causes and effects, where E and F are events that have causes and effects. But we can assess arguments for or against neuralism without a reduction of events to other entities. Indeed, such arguments should influence our views about the nature of events.

Lincoln was a tyrant', the referent of 'Lincoln' is the man Lincoln. On my view, the complementizer 'that' is a meaningful expression: the referent of 'that P' depends on the referent of 'that', which depends on the sense of 'P' (cf. Davidson 1968). This preserves the attractive features of Frege's view, while avoiding many objections to it. And remaining objections can be met, I contend, if one lets the sense of a referring expression depend on the context.

Since intentional explanations often involve 'that'-clauses, we want a clear view of what such clauses mean. But my interest in Fregean semantics is not limited to its clarificatory role. In Chapter 3, I extend the account of attitude ascription to explanation: the fact that P explains the fact that Q, iff the sense of 'P' explains the sense of 'Q'. This fits well with the common view that explanation is a relation between facts, if we follow Frege in taking facts to be the senses of true sentences. And given the event analysis, the senses of many true sentences can be construed as *ways of thinking about events*. The sense of 'Booth shot Lincoln', I claim, is a way of thinking about the event of Booth's shooting Lincoln.

If explanation often holds between ways of thinking about particular events, this invites the thought that causation (insofar as it is a relation between events) is the transitive *extensionalization* of explanation. To a first approximation: if C is the Φ -ish event, and D is the Ψ -ish event, and the fact that the Φ -ish event occurred explains why the Ψ -ish event occurred, then C caused D (no matter how these events are thought about, or described); and if C caused D, which caused E, then C caused E. Of course, not every explanation involves ways of thinking about events related as cause to effect. We must restrict attention to explanations *of an appropriate sort*. If 'appropriate' is a mere synonym for 'causal', the proposed sufficient condition may not be very interesting. But, I argue, one can avoid vicious circularity; and this bears on the question of whether our tryings and their mental causes are biochemical events. To see why, we need to step back and consider a central aspect of the mind-body problem.

2

Minds affect the world, at least in the sense that mental events cause non-mental events. Booth's decision to shoot Lincoln changed the course of history. In the summer of 1914, European leaders came to believe that a Bosnian Serb had killed the Austrian Archduke; these mental events had dramatic effects. Or to take a fictional example, there were causal consequences of Othello's coming to believe that Desdemona loved Cassio. Like Iago, most of us believe that affecting a person's thoughts will affect what that person does; and this belief influences our own actions when dealing

with others. Behaviour may not be the only somatic manifestation of the mental. For example, stress effects and placebo effects suggest that certain beliefs can affect a person's health. But the paradigmatic cases of mental causation are those in which a person's mental events cause her actions, which have a host of non-mental effects.

In my view, many bodily motions are already effects of actions. Shifting to a non-violent example, suppose Nora reached for her umbrella because she came to believe that it was raining. I think the acquisition of her belief caused an action/trying that caused a bodily motion, which caused the motion of the umbrella. Nonetheless, mental causes of actions can still be distal causes of bodily motions, which have non-mental effects. (I will not try to define 'mental', since our intuitions will be more secure than any definition. Nora's noticing the weather is a mental event, while the motion of her umbrella is not. If some mental events like pains are not intentional, so be it; though I am not concerned with such events.)

Indeed, it is a truism that bodily motions often have mental causes. But the truism is in tension with other plausible claims. Human bodily motions are typically caused by muscle contractions, which are caused by certain biochemical events that in turn have other biochemical causes, and so on—until one comes to events of sensory transduction, whose proximal causes lie outside the body in question. These facts appear to leave no room for mental causes. Correspondingly, they also make it seem that a bodily motion with mental causes would be overdetermined, like the death of a victim both shot and poisoned. But ordinary bodily motions should not be assimilated to paradigmatic cases of overdetermination. If I wave my hand, the motion of my hand does not seem to be the result of two independent (but coordinated) causal processes. This intuition is bolstered by counterfactual considerations. My hand would have remained motionless had relevant neuromuscular pathways been severed. So while there are clearly mental causes, it is not clear *how there can be* mental causes.²

One might think the mystery is easily solved: each mental cause of a human bodily motion *B* is a biochemical cause of *B*; or as it is sometimes put, our mental events just are certain neural events. Call this thesis 'neuralism'. (One could speak of 'biochemicalism'; but following philosophical

² I take as given, in light of Davidson (1963, 1967*b*, 1971) and more recent development of his arguments (e.g. Mele 1992), that there are no conceptual barriers to treating reasons as causes. One might deny that there are any mental causes; though if anyone but a philosopher denied having thoughts, or said that his thoughts never affected his body, this would be taken as a sign of madness. And in my view, evidence of mental causation abounds: people reach for umbrellas when it is raining; they drive to airports shortly before friends are due to arrive; they fill in circles on standardized tests; etc. (see Fodor 1975, 1987; Rey 1997). But my goal is not to convince sceptics that mental events have non-mental effects. I want to know how mental causation is possible, given facts that seem to exclude its possibility. I return to this point in Sect. 4 below.

custom, I will use ‘neural’ broadly to include changes in nervous systems that are not literally changes in neurons. For example, neuralism is compatible with the existence of mental causes that are—or are in part—changes in the level of hormones that affect action potentials.) Neuralism has many attractions. In particular, if mental causes *are* neural causes, there is nothing puzzling in the fact that certain bodily motions have mental *and* neural causes. Indeed, I understand neuralism in terms of the puzzle it is supposed to dispel. So for the most part, I will not distinguish between the following two claims: mental events are identical with certain neural events; and mental events are constituted by neural events, where ‘constitute’ is understood so that constituted events are not *causes distinct from* their constituting events.

Correspondingly, neuralism is true if mental events are fusions of biochemical events; and this is a point worth pausing over. Assume that we have an adequate grasp of which predicates count as mental and which as neural. This lets us speak of at least some mental events and neural events without settling questions of identity a priori. But neuralists need not say that mental events satisfy predicates like ‘has action potential p’. Put crudely, neuralism should be understood as a thesis that is compatible with mental events being relatively large and complex biochemical events. This makes appeal to mereology tempting, and perhaps unavoidable. So let us grant that some events are mereological fusions of other events, and that a fusion of neural events also counts as a neural event. In the context of debates about mental causation, this is to grant that if a part *p* of an event fusion *F* has an effect *e*, then *F* is not a cause of *e* distinct from *p*. That is, there is nothing puzzling in the fact that a fusion *and* some part(s) of the fusion caused a given effect. But this seems right. If we learned that some mental event was a fusion *F* of events that could be described in the language of neuroscience, there would be no further question about how some event caused by the parts of *F* (perhaps in coordination) could have a mental cause.³

This is *not* to say that neuralism is true if mental events are constituted by neural events in some way or other, with talk of mereological fusions being just one example of such a constitution relation. For it is crucial that event fusions are not *causes distinct from* their parts. And the mere fact that mental events are constituted by neural events in *some way or other* does not guarantee that mental events are not causes distinct from neural

³ Davidson (1970) claims that all causes are covered by strict physical laws. So on the assumption that strict laws are stated in terms of predicates satisfied by events involving *very* small objects, the following is a natural gloss of Davidson’s claim: C caused E only if C and E are both mereological sums of events, such that the parts of C are related to the parts of E by strict physical laws. (See Hornsby (1985), who argues that many event fusions will lack a kind of unity that genuine *causes* exhibit.)

events. There is a sense in which supervenient entities are constituted by subvenient entities; and the mental supervenes on the non-mental. But as we shall see, it does not follow that mental events are not causes distinct from non-mental events. And if one defines 'constitute' so that constituted events cannot be causes distinct from their constituting events, then event dualists will deny that mental events are constituted (in this demanding sense) by non-mental events.

On the other hand, talk of fusions may not capture the *only* notion of 'constitution' that will support neuralism. Perhaps there is an intermediately strong notion, such that constituted events are neither fusions of their constituting events nor causes distinct from their constituting events. But one cannot just assume that mental events bear some such metaphysical relation to events that satisfy untendentiously neural predicates. Absent an *account of how* constituted events can be causes that are not distinct from their constituting events, the main virtue of neuralism—viz. its simple explanation for why a bodily motion can have mental 'and' neural causes—has been lost. By way of comparison, suppose an event dualist appealed to an intermediately weak notion of being distinct, such that mental causes can be distinct from neural causes without overdetermining any effects of neural causes. One would want to hear more about this notion of distinctness; for therein lies the account of mental causation.

In short, one makes choices about where to do the hard work on the mind-body problem. Historically and theoretically, the following choice has been particularly important: should one adopt the simple account of mental causation, and embrace the consequences of saying that each human mental event is (or at least fails to be distinct from) some biochemical event; or should one eschew these consequences, and defend another conception of mental causation? From this perspective, to say that mental events *are* neural events is to say that mental events are related to certain neural events in a way that obviates the need for any further *account* (like the one I am proposing) of how an effect of those neural events can have mental causes. If mental events are identical with the relevant neural events, or fusions of them, no such account is needed. And one is free to characterize other notions of constitution, given which constituted events are not causes distinct from their constituting events. *Perhaps* mental events are constituted by neural events in some such sense. But the word 'constitution' is not a talisman that protects against objections to both neuralism and dualism; it labels work that needs doing.

One might claim that mental events are related to biochemical events, much as biochemical events are related to physical events; where 'physical' is interpreted in a demanding sense, characterized by reference to our best

theories in physics.⁴ But biochemical events may not be fusions of physical events. One cannot assume otherwise, claim that mental events are neural events in the same sense that neural events are physical events, and declare this an account of mental causation. Nor can one assume, without begging questions against event dualists, that mental events are just one more species of ‘macro’ event in a hierarchy of events that can in principle be described in physical terms. Even if biochemical events are constituted by (but not fusions of) physical events, defending this view may require a conception of causation like the one I will be urging here; and this conception might not be sufficiently motivated, without the independent support of reflection on mental causation. So one cannot make the puzzle of mental causation go away, simply by pointing to other macro causes. Mental causes may be special, at least epistemically, and perhaps ontologically. (I return to these issues in Chapters 5 through 7.)

Event dualists will say that mental causes fail to be biochemical causes, in any sense of ‘be’ that avoids the need for a substantive account of how an event can have mental causes *and* neural causes. There are reasons, to which I return, for adopting this view. That is, there are reasons for taking on the question of how a bodily motion can have distinct mental and neural causes, as opposed to the questions that neuralists must take on. But it is hard to see the force of objections to a view, absent any alternative; and it can seem that postulating non-neural mental causes must lead to disaster. Given that mental events are inner causes of bodily motions, it can seem that Cartesian dualism and neuralism exhaust the options; and Descartes’ picture of mental causation is unacceptable. Moreover, as Lewis (1966) and Armstrong (1968, 1970) note, the inner causes of bodily motions discovered by our best science are biochemical. So a challenge to neuralism requires at least some outline of a non-Cartesian alternative.

⁴ While our current theories are incomplete, and no doubt partly mistaken, assume that the standard model of particle physics is essentially correct: there are particles (and associated fields) of at least many of the posited kinds—quarks, electrons, photons, etc.; there is a stock of such particles (or a stock of more primitive entities to which quarks and such are reducible) out of which every material object in the universe is composed; and these fundamental particles have properties that are (or provide the reductive base for properties that are) at least importantly *like* the properties ascribed by the standard model, e.g. mass and charge. The fundamental particles and their interactions are governed by laws; and if these laws are not strict, at least they form a *closed* system, in that any apparent exceptions are explicable from within the system, i.e. by citing the fact that certain fundamental particles have certain fundamental properties. This provides at least a starting point for a characterization of physical events: any event involving only fundamental objects and properties, and any fusions of such events. It seems unlikely that any corrected version of the standard model will include reference to mentality. My aim is to defend a thesis which entails that mental events are not physical events in this demanding sense. (If ‘physical’ is characterized in a less demanding sense, e.g. as anything occurring in spacetime, event dualists may be able to grant that mental events are physical; see Crane and Mellor 1990.)

3

Returning now to my claim at the end of Section 1, suppose that: in general, event C caused event E if a fact suitably related to C explains (in an appropriate way) a fact suitably related to E; a fact suitably related to a particular event *c* explains a fact suitably related to a particular event *e*; and a fact suitably related to a third event *g* also explains a fact suitably related to *e*. Then *c* and *g* are both causes of *e*, since the occurrence of *e* (i.e. the fact that *e* occurred) is twice explained. But the relation between *c* and *g* is left open. If *e* is the motion of Booth's finger, the occurrence of *e* can be explained in neuroscientific terms, without recourse to the intentional idiom. Yet the occurrence of *e* can also be explained as the result of some action/trying, whose occurrence can be explained in overtly rationalizing terms. I will argue that *e* is not overdetermined in any objectionable sense, even if its mental causes are distinct from every neural cause of *e*, given that Booth's mental events (including his tryings) supervene on his neural events. And as we shall see, this supervenience thesis does not entail neuralism.

Elaborating this story requires work. In particular, it requires at least: a reply to overdetermination objections; a non-reductive account of *why* the mental supervenes on the non-mental; and a defence of the claims that causation can (without vicious circularity or implausibility) be characterized in terms of explanation. But the idea is not that mental causes fill gaps between certain neural causes of bodily motions. It is rather that the occurrence of a single event can be twice explained, even if the explanations do not share a common ontology of events. We cite causes in giving reasons for actions, and in giving impersonal explanations of various non-mental phenomena.⁵ The latter style of explanation can be applied to

⁵ I return to the distinction between rationalizing and impersonal explanations in Ch. 5. But let me say a few words here. A person's behaviour can often be revealed as a sensible course of action, given her goals and assumptions; while the behaviour of mere objects cannot be explained in this fashion. (I use 'behaviour' broadly; persons, plants, and protons all behave in various ways.) One might explain why Nora dropped a brick, by saying: Nora dropped the brick, because she wanted the brick to break, and believed the brick would break (when it hit the floor) if dropped. When dropped, the brick falls to the floor with a certain acceleration. One can explain why the brick falls as it does by citing the law of gravity. But the law does not reveal the brick's motion to be a sensible course of action, given certain goals and assumptions; the law tells us something about the arational workings of the world to which we belong. Similarly, events that occur inside a person's body can cause motions of that body, without the internal causes being reasons. Given any bodily motion B, one can in principle explain why B occurred without rationalizing anything. One can offer the same *kind* of impersonal explanation for why B occurred that one offers for why a brick falls to the floor, or for why a person shivers when cold: it happened because of external forces and/or internal causes that are not among the person's reasons for acting. But Nora's hand opened, because Nora wanted to drop the brick. This raises the question of whether rationalizing causes can be described in impersonal (e.g. biochemical) terms. To

persons—in the form of physics or neuroscience—since (necessarily) persons have bodies. But persons are individuals of a special sort; and differences between rationalizing and impersonal explanations may correspond to different styles of *individuating* causes of bodily motions. This may preclude the possibility that these explanations involve thinking about the same events in different ways. Instead of identifying mental events with neural events, we may have to recognize different (correct) ways of characterizing *as a series of events* what happens inside persons. The world affects us, and we affect it. Events inside us figure in causal chains that include events external to us. One can describe some such causal chains using impersonal terminology. But often one describes such chains in rationalizing terms. And rationalizing causes may not be describable in impersonal terms.

In Chapters 4 to 7, I argue that this picture offers a coherent and motivated alternative to neuralism, given the conclusions of Chapters 1 to 3. But while it is useful to frame the issue in terms of event identities, my claim is that advancing theses about the nature of mental events is *not* the best way to resolve puzzles about mental causation. We do better by critically examining our views about causation and explanation (cf. Burge (1989, 1992, 1993), Rudder-Baker (1995)). For example, a common view is that every genuine cause *C* has an effect *E*, such that the event pair $\langle C, E \rangle$ instantiates a conditional generalization of the right sort, where generalizations couched in intentional terms fail to be of this sort. Given some minimal assumptions, it follows that every intentional cause has a non-intentional description. Davidson (1970) advances a version of this view, requiring that singular causal claims be backed by *strict* laws of nature. It is unclear that every cause is covered by a law of any interesting sort. But in any case, many scientific explanations invoke laws that are hedged by *ceteris paribus* clauses.

There is a puzzle about how a hedged law can be non-vacuously true. But as I argue in Chapter 4, this puzzle can be resolved. Instead of imposing a necessary condition on causation, I offer a sufficient condition for explanation. Roughly, *F*₁ explains *F*₂ if *F*₁ is an instance of the fact that an event of type *T*₁ occurred; *F*₂ is an instance of the fact that an event of type *T*₂ occurred; and *ceteris paribus* if a *T*₁-event occurs, then a *T*₂-event occurs. The less rough proposal will exclude certain instances of *T*₁ and *T*₂ as relevant instances of the hedged law. And I argue that familiar objections to traditional covering-law models of explanation can be avoided, given my proposal about *ceteris paribus* laws. I also take the following to be a *ceteris paribus* law: if a person who wants Ψ to be the case comes to

use another Sellarsian metaphor: are causes visible in the manifest image visible in the scientific image?

believe that Φ -ing will make Ψ the case, then she will try to Φ .⁶ So the fact that someone acquired a certain belief can explain why she tried to do something. Similarly, the fact that she tried to do something can explain why her body moved. The fact that Nora's arm rose may well be explained by the fact that Nora raised her arm; where Nora's action (of raising her arm) *is* the event of Nora's trying to raise her arm. And, I suggest, tryings and bodily motions often instantiate *ceteris paribus* laws; if Nora tries to raise her arm, her arm typically rises. (This is not to say, however, that we always appeal to laws in *giving* intentional explanations.) In any case, suppose the onslaught of a belief explains the occurrence of a trying, and this explains why some bodily motion occurred. Then if causation is the extensionalization of explanation, mental events satisfy a sufficient condition for being causes *whether or not mental events are neural events*. This will let us see how acquiring a belief can cause a trying, and in turn a bodily motion, even if neuralism is false.

Let me note that the proposed sufficient condition for event causation is closely related to Fodor's (1989) account of why mental properties are causally relevant (despite being supervenient): a property T1 is causally relevant, if it is projected by a causal law, i.e. if for some property T2, it is a causal law that events with T1 are followed by events with T2; and there are (hedged) causal laws that project mental properties. Fodor goes on to suggest that cp laws have to be 'mediated by' physical mechanisms, while confessing to uncertainty about what this means. Given some such further claim, one can try to argue for Davidson's conclusion without his necessary condition on causation. For perhaps mental causes are covered by hedged laws, and only physical events are covered by hedged laws. But event dualists will challenge the latter assumption.⁷ And interestingly, Fodor (1989: 156) ends his paper as follows:

[I]f we *can't* get both the causal responsibility of the mental and an argument for physicalism, then it seems to me that we ought to give up the argument for physicalism. I'm not really convinced that it matters very much whether the mental is physical; still less that it matters very much whether we can prove that it is. Whereas, if it isn't literally true that my wanting is causally responsible for my reaching . . . then practically everything I believe about anything is false and it's the end of the world.

⁶ This is a simplification. The person might *intend* to Φ when some condition C obtains. But then the person will try to Ψ , *ceteris paribus*, if she comes to believe that C obtains (see Ch. 3).

⁷ As Fodor (1989: 159, n. 18) says, if you want to get physicalism out, you have to put physicalism in—say, in the form of 'the independent assumption that the mechanism of intentional causation is physical'. Segal and Sober (1991) challenge Fodor's account; but see Sect. 4.2 of Ch. 5 below.

Cataclysmic rhetoric aside, I agree with this ordering of priorities; and one can substitute ‘neural’ for ‘physical’. Indeed, one might see my account of mental causation as an attempt to develop Fodor’s proposal as part of a view according to which neuralism *undermines* the idea that we do things for reasons. But one cannot reject neuralism, maintain that there are mentalistic *ceteris paribus* laws, and then just assume that pairs of events covered by *ceteris paribus* laws are related as cause to effect. Defending this assumption is where I think the hard work lies; though I think the needed work can be done.

With this possible account of mental causation in hand, one can better appreciate the various reasons for rejecting neuralism. Given an alternative, a weighty burden of proof lies with those who say that intentional explanations cite the same causes (thought about in different ways) as certain impersonal explanations. I think neuralism actually threatens the idea that reasons are causes by identifying mental events with events characterizable without reference to *persons*. Identifying reasons with impersonally characterized events threatens our view of ourselves as agents whose actions are *free*. And if mind–brain identity theories repeatedly invite charges of epiphenomenalism (see Heil and Mele 1993), this suggests that such theories have not really helped to show how minds can affect the world.

In Chapter 5, I also argue against neuralism more directly. Following Hornsby (1981) and others, I contend that while mental events are spatially located, their location is typically less determinate than that of any biochemical event. This reinforces a claim I hinted at above: intentional explanations individuate causes in a way that makes mental events poor candidates for identification with events not individuated in rationalizing terms. Neuralism also has the implausible consequence that certain biochemical events have propositional contents. (Othello’s coming to believe that *Desdemona loved Cassio* was an event with the content expressed by the italicized sentence; see Chapter 3. So if this mental episode was some biochemical event B, then B had the same content.) In the appendix, I focus on how this saddles neuralists with the unenviable task of explaining how a biochemical event *could* be semantically evaluable in this sense; whereas event dualists can and should avoid this project, which turns out to be the ambitiously reductionistic project of providing a so-called naturalistic theory of content. (If we think of the mind–body problem as having causation and content strands, adopting neuralism in response to the former makes responding to the latter harder—and perhaps impossible.)

4

Often, disagreement in philosophy concerns what one should be trying to do. So let me end this introduction with a few remarks about what I am

trying to do, and why. Given the number of topics I connect—action sentences, Fregean thoughts, *ceteris paribus* laws, and so on—it will come as no surprise that I agree with Sellars (1963): the aim of philosophy is ‘to understand how things in the broadest possible sense of the term hang together in the broadest possible sense of the term’. But time is short, and so we don’t discuss just *any* things. (We leave out talk of sealing wax, of cabbages, and kings.) Motivation for philosophical inquiry is usually rooted in *puzzles*: questions for which every imaginable answer seems wrong, especially when such questions concern the nature of persons, whose nature it is to think. And certain facts come to seem less puzzling, I claim, given the proposed conception of how the topics under discussion hang together.

In particular, the proposal lets us see how mental events can cause bodily motions, even given that such motions have neural causes. This is not the only puzzling fact about thinkers. In the appendix, I briefly address the related question of how error is possible, given that beliefs are in some sense representations of the environment. (How can a thinker *represent* that which is not the case? Yet if error is impossible, it seems that there is no interesting notion of mental content.) And the hardest questions about minds are not about causation or content. But I leave the topic of consciousness to braver souls, hoping that it is (at least largely) independent of the issues addressed here. Providing, motivating, and defending an alternative to neuralism is hard enough although a successful defence of event dualism as an account of mental causation may have implications elsewhere in the philosophy of mind.

It is worth being explicitly clear, however, about the *kind* of question to which my proposal is a purported answer. Nozick (1981) provides a model of the kind of project I have in mind. So let me set mental causation aside, for the moment, to consider why one might want a theory of knowledge. A central puzzle of epistemology is that we know anything, given facts of the sort adduced by sceptics. You know that you are currently reading a page of text. But certain modal facts appear to exclude the possibility of such knowledge. Things would seem as they do now if you were being deceived by a clever demon or a nefarious neurosurgeon (who envatted your brain a short while ago). You don’t know that you’re not being deceived. So how can you know that you are reading a page of text?

This problem has a familiar form: it is the case that P; but it is also the case that Q; and it is hard to see *how it could be the case* that P, given that Q. One response to such problems is to deny P and/or its apparent excluder Q. But it strikes me as obvious that we often know things, despite our fallibility. Like Moore (1925), I see no prospect for a convincing proof that knowledge is impossible based on premisses connecting fallibility to the impossibility of knowledge. For I don’t see how such premisses could

be more compelling than ordinary knowledge claims. And any valid sceptical argument can be rephrased as a *reductio* of some tendentious epistemic thesis. Similarly, it is hard to imagine premisses more compelling than (but which entail the denial of) the claim that people are fallible, even if recognizing our fallibility makes it hard to see how knowledge is possible.⁸

It is equally unlikely that one can establish the possibility of knowledge, using only premisses a sceptic would accept. But if one is convinced that knowledge is possible, one doesn't need a proof to that effect, unless one is trying to convince someone else. If the goal is to understand how P, one need not *make* anyone believe that P. (If some people deny that anyone knows anything, so be it; one is not obliged to convert the infidels.) Moreover, the existence of actual sceptics is irrelevant. Sceptical arguments would not cease to be bothersome if it were discovered that the alleged sceptics were joking, or that it is psychologically impossible to maintain scepticism. An argument can reveal a tension *in one's own thinking*, even if everyone agrees that the conclusion is false. By my own lights, deceivable people know things. An argument can show me that I do not see how my views on this score can be compatible.

Put another way, I may be *inclined* to think that sceptical possibilities exclude knowledge, even if I know otherwise. A good response to scepticism will address this inclination, by making explicit some tacit assumptions about knowledge that I judge (all things considered) to be

⁸ Tensions can take the form of an outright contradiction between seemingly obvious propositions. But more often one has the sense that P and Q do not cohere, even though one cannot yet formulate the trouble-making background assumptions. Still, some people may see no puzzle, even after informed reflection. (Consider how most of us would react to the question of how water can flow downhill, given that steam rises—or how fish can swim, given that turnips can't. What assumptions would lead one to see *problems* here? In the former case, perhaps mistaken views that a course in chemistry would correct; in the latter case, who knows?) It is hard to argue that one *ought* to be puzzled by something: where I see a paradox, you may see an obvious mistake. And one cannot demand that others see a tension, so they can be shown how to resolve it. But even if the rest of us cannot *make* others feel certain tensions, it does not follow that *our* predicament stems from an irrational overindulgence of metaphysics. I emphasize this point, because my views here are in line with those of Burge (1993) and Rudder-Baker (1995), who urge that we emphasize our confidence in mental explanation over metaphysical principles that make mental causation into a mystery. And fairly or not, these philosophers have been charged with adopting the following overly relaxed attitude: it would be silly to deny, on metaphysical grounds, that mental events have effects; so don't worry, be happy (see Kim 1995). Antony (1995: 160) lampoons those who 'see no problem about reconciling our folksy conviction that what we *think* matters to what we *do*, with our more tutored views about the structure of reality and the nature of causation'; and this attitude is equated with offering 'deflationary' responses to 'philosophical puzzles about mental causation', by emphasizing explanation instead of causation. Perhaps this makes me a deflationist. But to offer a response is already to recognize the puzzles. And defending deflationism requires effort that the unworried would never expend. I return to these issues in Ch. 7.

mistaken—even though I am favourably disposed towards these assumptions, when they remain unstated (or when I do not focus on their implications). It may be a piece of common sense *that* fallibility does not exclude knowledge. But to show oneself *why* fallibility does not exclude knowledge, one needs to show that one's inclination to think otherwise is rooted in a conception of knowledge that is partly misguided (and misleading). So a good account of knowledge will make it clear how fallible agents can know things, by revealing aspects of our conception of knowledge that dispose us towards puzzlement. This will remove a reason for thinking that knowledge is impossible, and in that sense explain how knowledge is possible.

Intellectual tensions are themselves puzzling things. How can one fail to see how P is possible (given an apparent excluder Q) if one sees that P is actual? Nonetheless, we sometimes experience such tensions; one's beliefs can exhibit a failure to 'hang together' that is literally queasy-making, when the lack of integration is noticed. This phenomenon is not peculiar to philosophy, or even to science broadly construed. Novelty stores often sell toys that pose the challenge of showing how something actual is physically possible. (The ring can be separated from the post, but how?) And I assume that removing tensions in one's thought is a valuable activity, once we turn from toys to puzzles concerning the nature of persons: how can our actions be free (and morally evaluable), given a world governed by natural laws; how is experience possible, given that subjects exist in a world composed of physical objects; how is knowledge possible, given that knowers are fallible?

Like Nozick, I think philosophical theories are best viewed as attempts to clarify and answer such questions.⁹ Theories of knowledge should be assessed, in part, by how well they help resolve the tension revealed by scepticism. Similarly, if philosophers say that mental events are (not) biochemical events, I think such claims should be assessed according to how well they help resolve the tensions associated with mental causation. The question is not *whether* mental events have non-mental effects, or *whether*

⁹ Wittgensteinians will say that *theorizing* is the wrong response to such questions; and 'theorizing' may be the wrong word for the kind of response I have in mind. In any case, I try to defend some claims that help show why certain facts about the mental are compatible with apparently conflicting facts. Perhaps philosophical theses are best viewed as tools for escaping from confusions, not as straightforwardly factual claims (like 'The cat is on the mat'). But one can say, to borrow a Quinean metaphor, that each portion of a person's cognitive web is meaningful by virtue of being part of a system of beliefs that serves as the person's overall conception of her world. And recognition that thesis T has helped one resolve puzzles may well be a good reason for retaining T in one's cognitive web, though revision to such aspects of one's web may be driven more by 'housekeeping' considerations than by observation (cf. Carnap's (1950) suggestion that, in one sense, metaphysical questions call for pragmatic decisions about how to speak—and how to carve up the work).

we have false beliefs. (They do.) The question is *how there can be* mental causation, given various facts about persons. But the form of such questions at least suggests the form of helpful answers.

Faced with a tension between P and an apparent excluder Q, an obvious strategy is to search for a condition C, such that C is sufficient for P; one can see how C could obtain, given Q; and seeing that C is sufficient for P helps one see that some tempting assumptions (which entail the incompatibility of P and Q) are false. From this perspective, one wants a sufficient condition for knowledge, such that: one can see how this condition can obtain, even given sceptical possibilities; and stating this condition helps one expose and reject tempting claims that would render knowledge incompatible with fallibility. Similarly, I want a sufficient condition for being a cause of a bodily motion, such that: one can see how mental events can satisfy this condition, even though all bodily motions have neural causes; and the resulting proposal helps us see why we found mental causation to be puzzling in the first place.

In the case of epistemology, many theorists have urged us to abandon the assumption that knowledge is closed under entailment. For then one can grant that a subject S can fail to know that Q (e.g. that she is not being deceived), even if S knows that P and that P entails Q. So one might look for a theory—a sufficient condition for having knowledge—according to which knowledge is not closed under entailment.¹⁰ Sceptics are unlikely to accept any such theory. But that does not matter if the goal is not to convince sceptics. Similarly, those who deny the existence of mental causes are unlikely to accept my proposed sufficient condition for mental causation. But if the aim is to resolve a tension for those who grant that there are mental causes, then what matters is whether that audience finds the proposal plausible and helpful in resolving the tension. Discovering such a theory is hard enough. For given my pretheoretic intuitions, I will find many theories implausible and/or unhelpful.

That said, I may (and often should) revise my intuitions in light of an otherwise plausible theory. Not *all* of my pretheoretic judgements about what counts as knowledge (right action, causation, etc.) will conform to any simple theory; and as discussions of reflective equilibrium suggest, there is epistemic value in having judgements such that one can provide a relatively simple systematization of them. More importantly, it would be absurd to deny that the value of resolving puzzles ever justifies revision to one's judgements, if such judgements are manifestations of an overall conception that (by one's own lights) contains some internal tensions. Still, I cannot always give up an intuition, just because doing so would let me

¹⁰ See Dretske 1970; Goldman 1986. Nozick discusses this point at length. Perhaps one can argue that knowledge is closed under entailment after all; see Lewis (1997). But that is another issue.

achieve cognitive harmony. That is why intellectual tensions are persistent. Nor should one be cavalier in rejecting pretheoretic views. The goal is to see how one's views fit together, perhaps with minor revision. So it is self-defeating to reject intuitions too quickly. I want not just any coherent set of beliefs, but a coherent set of beliefs that is recognizably *mine*.

On this view, philosophical theorizing is not an attempt to define terms (like 'know' and 'cause'), or to systematize intuitions. Nor is the goal to provide a theory of concepts if these are taken to be mental representations that thinkers typically use in categorization. The aim is to resolve intellectual tensions. A person's intuitions will *constrain* which theories she will and should find to be acceptable ways of resolving tensions. Reflective thinkers will and should reject some sufficient conditions for knowledge (or mental causation) as too revisionary.¹¹ Moreover, in emphasizing what Nozick calls 'domestic' rather than 'foreign' policy—fixing one's own beliefs, as opposed to proselytizing—one need not deny the relevance of others' intuitions. If my intuitions are idiosyncratic, I should take seriously the possibility that my intuitions are distorted. One wants to address tensions that others also experience, in a way that others would also find plausible. Finally, to emphasize the role of puzzles is not to deny that one aims at truth in philosophy. Resolving tensions in one's overall conception of the world is part of making one's beliefs fit the facts.

Once again, a philosophical account of mental causation should provide a sufficient condition for being a mental cause of a bodily motion, such that this condition helps us see how bodily motions have both mental and neural causes. (That is, a good theory should provide at least this much.) My proposal combines event dualism and a conception of causation as the transitive extensionalization of explanation. I argue that this is preferable, all things considered, to neuralism. A crucial component of my view is that actions typically cause bodily motions. So let me now turn to the argument for this claim.

¹¹ A theory can tell us something about our concept of knowledge (and what knowledge is). But intuitions are not best viewed as *data*, in the way that speakers' intuitions serve as data for theories in linguistics; although if a theory relieves a tension for individuals who felt it—and not because the theory has some arational soothing effect—that tells in favour of the theory. Stich and Laurence (1994) rightly ask what philosophers are doing when offering theories, such that intuitions are germane to theory choice, but the goal is not to describe intuitions (or reveal an innate competence *à la Chomsky*). But solving puzzles can be an intellectually respectable activity; see e.g. Kuhn (1970). Note that *necessary* conditions for knowledge (causation, etc.) are not required; the issue will be whether some sufficient condition is plausible (and plausibly satisfied in relevant cases). So a philosophical theory need be a traditional *analysis* of anything.