

OUP Copyright

## CHAPTER 26

# Agency, causation, and freedom

### Chapter contents

- Session 1 **Agency, freedom, and neuropsychiatry** 718
- Session 2 **Agency and volitions** 720
- Session 3 **Arguments for a causal theory of mind** 724
- Session 4 **Event causation, agent causation, and irrationality** 727
- Session 5 **A non-causal account of agency?** 731
- Session 6 **Freedom and determinacy** 732
- Reading guide** 735
- References** 735

OUP Copyright

## Introduction

One recurrent theme of this book is the relation between reasons and causes in psychiatry. Whereas many disciplines operate with one or other form of explanation and interpretation, psychiatry operates with both. Nowhere is the tension between the two as great as the issue of freedom versus causal determinism.

Giving, asking for, and acting on reasons appears to presuppose freedom. Summarizing Kant, McDowell (1994, p. 5) puts this connection like this: 'rational necessitation is not just compatible with freedom but constitutive of it. In a slogan, the space of reasons is the realm of freedom'.

But, as the first session of this chapter will outline, the success of recent brain imaging techniques looks to cast doubt on the possibility of such freedom. In other words, our increasing ability to explain neurological function in causal terms seems to put at risk the very possibility of freedom on which reasons depend.

This chapter will examine the connection between agency, causation, and freedom. The initial route into the area will be by consideration of some issues raised by research using brain imaging techniques into disorders of agency. The first reading (linked with Exercise 1—Spence, 1996a), which concerns the exact timing of mental and physical events, sheds light both on specific disorders of agency such as 'made actions' and thought insertion but may also shed light more generally on the very idea of the freedom of the will.

To put this in a broader context, there has been considerable and inconclusive philosophical debate about whether freedom—which is central to the idea of a responsible agent—really is compatible with our growing knowledge of the workings of the physical world at least since Hobbes' time. The basic issues of that debate are almost paradoxically easy to understand and yet give rise to no clear solutions. Assume for a moment that every event in the history of the world is the result of a prior cause. This is to assume the truth of 'determinism'. If so, then whenever I think that I have a free choice of action, since the action I do choose is an event in the history of the world, it is thus caused by prior events. One may want to say that it was always going to happen. And this seems to cast doubt on my freedom to choose.

Of course most scientists now believe that determinism is false at least at the level of the microscopic. (There is *no* causal explanation of why a particular radioactive particle decays at the moment it does. Radioactive decay is *indeterministic*.) However, this does not seem to make free will any the less problematic. Rather than being caused, our actions may be the result of indeterministic whimsy at the quantum level!

## Agency: a more tractable issue?

Although we will return to the issue of freedom at the very end of the chapter, we will focus on a more tractable issue raised by the clinical research into made actions, thought insertion and such like. What is it to be the agent of an action, whether mental or physical? To approach this question we will set out some of the background to the debate on whether reasons are causes. Can the

very same mental states, which are invoked to explain and justify speech and action, also play a *causal* role in generating that speech and action? Are mental states part of a mechanical mind, a system of causal pushes and pulls? Or are they instead part of a different kind of description, perhaps abstract elements in a broader calculus of action explanation? What connects this issue with that of freedom is that rather than running shy of causes, a number of philosophers have argued that the very idea of action requires that it is part of the causal order.

## Agency and the mind-body problem

Different answers to the question of the relation between action and causation separate different positions taken up in the mind-body debate, which was discussed in Chapter 23. Functionalism, for example, holds that mental states are individuated by their causal roles or functions. Mental states are *caused* by perceptions and other mental states and *cause* actions and other mental states. Dennett's Intentional Stance, by contrast, holds that mental states are *abstracta* like centres of gravity, which are invoked for making predictions about behaviour but are not themselves causally active. (Dennett does think that there are causal accounts that are relevant to action, such as those provided by neurophysiology. But these do not talk of *mental* types.)

## Agency, causes, and reasons

This chapter will approach the reasons-causes debate through the central issue of whether causality can be used in the analysis of agency as an answer to the question: What is the difference between action and mere movement? Roughly, a causal theory of action attempts to spell out this difference by saying that actions but not movements are caused by reasons. Thus what has to be added to a mere movement to make it an action is the fact that it was caused by beliefs and desires. So, for example, the mere movement of a hand in the direction of a coffee cup, is the *action* of *reaching for the cup* if it is caused by appropriate mental states. (These may be the desire to drink and the belief that by reaching for the cup one will be able to drink from it.) In fact, however, we will see that it has to be caused *in an appropriate way*. So psychopathological cases of 'made' actions—which result from reasons that the agent disowns—present interesting material for philosophical reflection.

A causal theory of action is an analytic project, aimed at shedding light on the notion of agency (of who or what are agents, of when a movement is an action and so on) by consideration of causality. But aside from that analytic project, the issue of whether mental states can have causal effects is also important for understanding the place of mind in nature as has just been suggested. One common guiding thought is that mental properties cannot be *real* properties, genuine parts of nature, if they cannot stand in such relations. So if this is disputed it will have important consequences for how we understand mental properties by comparison with other properties charted by the physical sciences and thus the status of psychology and psychiatry.

### The plan of the chapter

- ◆ *Session 1* will examine agency and freedom from a neurological perspective.
- ◆ *Session 2* will examine non-causal philosophical accounts of agency from the 1960s.
- ◆ *Session 3* will introduce Donald Davidson's influential causal theory.
- ◆ *Session 4* will consider in more detail just how a causal theory is supposed to work and consider whether it can shed light on irrational action.
- ◆ *Session 5* will examine recent criticism of a causal theory.
- ◆ *Session 6* will return to the broader issue of the freedom presupposed by the notion of agency.

### Session 1 Agency, freedom, and neuropsychiatry

The first reading attempts to draw conclusions for the nature both of agency and free will from both accounts of psychopathological symptoms and also neurological work. It is a challenging and very clear research paper by Sean Spence, a contemporary UK neurophysiologist specializing in brain imaging.

#### EXERCISE 1

(45 minutes)

Read the two extracts from:

Spence, S.A. (1996a). Free will in the light of neuropsychiatry. *Philosophy Psychiatry, & Psychology*, 3: pp. 75–90. (Extracts: pp. 78–81, 83–85)

Link with Reading 26.1

Take note of the brief clinical vignettes describing made action, thought insertion, and such like.

- ◆ What do these suggest about the nature of agency and freedom?
- ◆ How do these relate to the description of Libet's work on the timing of mental and neuronal events?
- ◆ What are the novel consequences of such brain imaging work for free will?

#### Neuropsychiatry and free will

The paper begins by outlining the philosophical problem of free will: that there is a conflict between our belief that we can freely choose how to act and our belief that what happens in the world, including our actions, is the result of antecedent causes. (Spence suggests that this way of putting the problem makes free will *merely* a belief. But talk of belief in the quotation could be replaced by 'putative fact'.)

Without further details, one can already begin to see the sort of problem that the paper might address, given its title. Given that research in neuropsychiatry will, no doubt, increasingly reveal the causal processes that 'underlie' mental activity, that research will increasingly focus attention on the already familiar philosophical paradox. (To say that physical processes 'underlie' mental processes is already to take up a stance in the mind–body problem. But it is at least widely accepted that possession of brain states is causally necessary to have mental states, at least for humans.)

#### The two strands of argument

In fact, Spence goes on to make two different kinds of more specific claim. One concerns what might be called the 'phenomenology' of agency. It is here that the case vignettes have their role. These imply that there is some continuity between non-pathological experiences of thoughts popping into one's mind and pathological experiences of made thoughts and made actions. The other concerns experiments based on work by Libet, on the relative timing of mental and physical phenomena. These suggest that conscious phenomena lag behind characteristic neurophysical precursors.

#### Experiences of altered volition

To take these two ideas in turn. There are perfectly normal cases of what Spence terms 'altered volition'. These are cases where, unlike a normal experience of deliberately saying or thinking something, one either blurts something out or a thought pops unbidden into one's mind. In these cases, the 'experience of voluntary action is diminished or lost'. Subjects 'do not experience as freely willed action the impulses which produced these acts' (p. 78).

Spence goes on to consider two kinds of pathological case: the alien hand syndrome and passivity phenomena. Having set out case examples of alien hand syndrome, he summarizes them thus:

As may be seen from these accounts the alien hand is experienced as behaving in an autonomous and purposeful way. Its interference with normal activity is such as to provoke patients to developing strategies to restrain or 'distract' the hand. The patients acknowledge that the limb belongs to them but that its actions are not under their control. Yet they do not attribute its activity to outside forces (cf passivity) . . . The subject loses the *experience* of free will (with respect to the affected limb), and develops an accommodating abnormal experiential belief that the limb is autonomous. (pp. 80–81)

Spence then turns to passivity phenomena: made thoughts, made actions, made affects, and made impulses, all of which are symptoms of schizophrenia but also found in other diseases. 'The subject experiences their thoughts, actions feelings and drives as having been influenced or replaced by those of external agencies' (p. 81). Spence suggests that these also are best understood as resulting from abnormal experiences of agency or will. Drawing on work by Frith, he suggests continuity between the idea that such subjects are abnormally 'monitoring' their own volitional processes (thoughts and actions) and that they might be *experiencing* their agency abnormally. In some cases subjects do not

experience a sense of volition, despite acts occurring that appear purposeful. In others the act is experienced as volitional but resulting from an impulse that is not experienced as willed by its subject.

There are differences between these two sorts of pathological case and normal cases of a lack of experience of volition—central is the delusional elaboration of an outside locus of agency—but there is also continuity. Spence concludes that ‘seemingly “purposeful” action and “insightful” thought may be... distinct from the sense of “will” or “ownership” which normally attends them’ (p. 83).

### Summing up the first strand

So this is one strand of the paper. Although we might think of exercises of free will in action and thought as always conscious, there are both normal and pathological instances where we have thoughts or we act apparently deliberately, but do not experience those acts or thoughts as consciously willed by us. This raises an interesting pair of questions: what is to *experience* thought or action as willed by oneself? and what is it for a thought or action to *be* willed by oneself? Light will be shed on these throughout the chapter. But first we will consider the second strand of Spence’s work.

## The second strand of Spence’s argument

### The connection between the two strands

In the second strand of the paper, Spence looks to the connection between the experience of freely willed action and neuronal events. In fact this separation into two strands is a little artificial. In the passage quoted above, Spence actually says: ‘seemingly “purposeful” action and “insightful” thought may be *neurally* distinct from the sense of “will” or “ownership” which normally attends them’. But at that stage in the paper, it is not clear what justification there is for such a qualification (as opposed to saying simply that there is an experiential or conceptual distinctness).

### Brain imaging results

The empirical work behind the second strand is Libet’s work on the timing of mental and neuronal events. Subjects were asked to note when, according to a spatial clock they could see, they chose to make a decision to move part of their body. Meanwhile their EEG and EMG data were recorded. Libet noted that there was a characteristic electrical change 850 milliseconds prior to the actual movement but also 350 ms prior to the *subjective experience* of deciding to move. ‘Thus Libet *et al.* conclude that volitional activity is initiated non-consciously’ (p. 83).

This seems a radical conclusion. It suggests that what seemed like a free decision (that one will choose *now!* to move one’s finger) is precipitated by a prior non-conscious neuronal event, 350 ms earlier. And this seems to imply that the conscious decision was not itself free. As Spence notes, there is no chance of preserving a role for conscious free choice as a *veto* for an action already initiated non-consciously because that conscious event would also have to be pre-empted by a suitable neuronal event. So

if there is free will, it happens at the level of non-conscious processing and not at a conscious level of act or thought initiation.

### What if 350 ms were 35 seconds?

The time-lag described by Spence is very short and this can disguise the radical nature of his claims. One way to bring out the threat to free will is to imagine that the 350 ms were, say 3.5 seconds or even 35 seconds.

#### EXERCISE 2

(15 minutes)

Before reading on, think what would be the consequences if the time-lag between the key measured brain event and the perceived experience of willing were much longer? Imagine yourself to be the experimental subject. What would happen if you were told when such an event had been detected? Could you change your mind?

### Thirty-five seconds and freedom

If, unknown to a subject—you, say—Libet can predict 35 seconds before you make a decision that you will make it at a particular time, it does not seem that you are really free. You are a puppet of your neuronal events. To bring out the paradoxicality of this, now imagine that within this 35 second window, say 10 seconds before you are due to make your decision, Libet tells you exactly when you are going to make it. You cannot *now* stop making it in 10 seconds time, because that change of mind would require that there had already been another neuronal event 35 seconds before it (i.e. 10 seconds before the measured neuronal event). In fact if that new first event *had* already occurred, Libet would not have told you that you were going to make the decision as he originally predicted it. But if so, you would not now have changed your mind and so the new first event would not have occurred.

Such directly paradoxical possibilities are ruled out in the actual case by the small time interval actually found. There simply would not be time to react. But the fact we are saved from such counter-intuitive conclusions merely by the contingency of the time-scale should be small comfort.

### Are the empirical findings secure?

We will return to the issues this paper raises for freedom of the will at the end of the chapter. But it is worth flagging two considerations. One is an empirical matter. Spence’s conclusions could be resisted, and the subjective conscious experience of decision-initiation could be equated with the apparently earlier neuronal event if there was an equal lag between looking at the clock and noting its time. If so, the apparent lag between the neuronal events that initiate action and the conscious experience could be explained away as the result of time taken subsequently to note the position of the clock.

### A philosophical response

The other is a broader philosophical matter. Spence says that ‘the reductionist, materialist perspective underlying the current

research has been “pushed” to its logical conclusion. Consciousness is not the initiator of willed action’ (p. 88). But in fact the diagrams of conscious experience lagging behind neuronal events looks more like an ‘epiphenomenalist’ position in which the mind is somehow the result of neuronal processes but separate from them. A materialist, reductionist will want to know just *what* the conscious processes are processes *of*. If a person changes mentally in the instant that he or she experiences a conscious decision (the result of a prior neuronal event), then surely there must be some *physical* change at the same time that explains that mental change. If not, it looks as though the mind really is separate from the brain. So a materialist would expect there to be further neuronal events, after the initial events found by Libet and coinciding with the change in conscious experience.

Let us suppose that there are such events and that they are caused by events 350 ms before them (the events identified by Libet). Does this undermine free will? It seems that it does not add anything further to the problem that already existed. If the brain is a deterministic machine then its states stand in potentially predictable causal relations. Add in the materialistic assumption that brain states somehow determine mental states and this alone gives the picture just outlined. For any mental state, there are simultaneous brain states that determine it, and these in turn are caused by prior brain states according to yet-to-be-discovered neurological laws.

What this suggests is that the problem of free will requires thinking about the relation between freedom and agency, on the one hand, and causal laws on the other (as in fact Spence says in his conclusion). So to shed light on that we will now turn to the nature of agency and its relation to causality. We will focus on the question of whether the reasons we have for acting in paradigmatic exercises of free will are also the causes of action. Are reasons causes? As will be illustrated, the current orthodoxy is that far from being in conflict, the very idea of agency is tied to that of causality. An action is caused by mental states.

### Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. What do the case vignettes in Spence’s paper suggest about the experience of free will? Is there a unified experience?
2. Does Libet’s experiment really undermine free will? What interpretations could be offered, according to Spence?
3. What difference would imagining a greater time lag between prior brain event and conscious decision make?
4. Is Spence’s interpretation of Libet a triumph of materialism, materialist analysis taken to the limit?

## Session 2 Agency and volitions

### Non-causal accounts in the 1950s

The first session outlined the apparent conflict between some recent experimental work on the timing of neural events and our everyday understanding of free will. This session will look at some influential arguments that aim to show that action explanation has nothing to do with causation. Historically this approach was promoted in the 1950s by a generation of philosophers of mind such as Gilbert Ryle (1900–76) and the American A.I. Melden (1910–91), working very much in the tradition of linguistic philosophy. They argue that understanding action is of a very different form than explaining events causally and thus stress the difference between the human and the natural sciences.

The current prevalence of causal approaches to the analysis of action was provoked by Donald Davidson’s attack in 1969 on that previous non-causal orthodoxy. More recently, the current causal orthodoxy has itself come under fire again from non-causalists. So we will begin with the first generation of non-causalists. This will sharpen the conflict between the causal explanation of events that will increasingly be a feature of neurological science.

### Cartesianism and causality

Ryle’s work has already come under discussion in Chapter 22. Ryle was a key figure in the version of Anglo-American philosophy sometimes called Oxford or Ordinary Language Philosophy. We will examine a chapter from his attack on Cartesian philosophy of mind: *The Concept of Mind* (1963). In the course of this, Ryle attacks a causal construal of the role of reasons. Ryle’s target here is what may seem the rather strange combination of Cartesian dualism and immaterialism. However, Ryle’s arguments were also taken to count against materialist causal positions.

## Ryle’s critique of the Myth of Volitions

### EXERCISE 3

(30 minutes)

Read the extract from:

Ryle, G. (1963). *The Concept of Mind*. London: Penguin, pp. 61–80. (Extract: pp. 62–66)

Link with Reading 26.2

Isolate and assess Ryle’s arguments against the Myth of Volitions. What account of action does he propose in its place? Think also whether Ryle’s arguments would count against modern materialist causal theories of action such as functionalism.

### The Myth of Volitions

The chapter of the *Concept of Mind* concerned with the will is a major plank of Ryle’s argument in that book. In that chapter, Ryle aims to destroy a central element of the Cartesian picture of

mind. The object of Ryle's fire is a philosophical theory of the will, construed as an 'executive' mental organ. It is this organ that mediates between thoughts or wishes that are 'unexecutive' and actions. The mental intermediaries that need to be added to a wish, for example, to generate action are *volitions*. These are special internal mental events and, Ryle suggests, are more specifically a special kind of mental action. (That they are really internal mental actions underpins one of Ryle's arguments against them.) It is the role of the faculty of will to generate these volitions. Hence Ryle's description of this theory as the 'Myth of Volitions'.

But it is worth focusing first on the problem for which the volitions were intended to be the solution. On a Cartesian world view, there are two possible explanations for changes in the motion of inanimate objects or for the movements of animate bodies. They can either be brought about by the motion of other matter. Or in the case of human beings they can result from 'thrusts of another kind' (p. 62). This other internal source for movement is the will. This general distinction corresponds to, and is meant to explain, the distinction between mere movements and actions. An action, on this theory, is a movement brought about in the right way: as a result of an inner mental episode.

#### A hint of a causal theory

Although this is not particularly emphasized by Ryle, he suggests that the relation between a volition and the movement (and thus the action) it produces is a *causal* relation.

And so to say that a person pulled the trigger intentionally is to express at least a conjunctive proposition, asserting the occurrence of one act on the physical stage and another on the mental stage; and according to most versions of the myth, it is to express a *causal* proposition, asserting that the bodily act of pulling the trigger was the effect of the mental act of willing to pull the trigger. (p. 62)

Contemporary causal theories of action continue with this last suggestion—despite Ryle's criticisms here—but outside the context of a specifically Cartesian conception of mind. One aim of a causal theory of mind is to explain the difference between deliberate actions and mere (involuntary) movements by specifying the defining difference in the causal aetiology of actions and mere movements. Thus one of the issues in assessing Ryle's attack on the myth of volitions is the extent to which it is successfully directed towards any causal theory of the mind or whether it turns on a Cartesian version of that kind of theory.

#### Four criticisms

Ryle offers four arguments against the Myth of Volitions. They are based on:

1. the lack of empirical evidence for volitions;
2. the impossibility of third person epistemology;
3. the inexplicable connection between volitions and actions;
4. a dilemma about their status as actions.

#### No evidence

Ryle offers a number of loosely related objections that are designed to undermine the general plausibility of the theory of volitions. He observes that contrary to what the theory would lead one to expect, no one ever offers descriptions of their own volitions. They never report undertaking slow or difficult ones, for example. Nor would they know what to say about them if pressed because the descriptions that apply to other actions seem to get no grip with volitions.

Ryle summarizes this objection in the following way. Volitions are never ordinarily reported; they lack non-academic names; there are no principles to determine their frequency, duration, or strength. Ryle concludes from these considerations that there is no empirical evidence for the existence of volitions or the will so construed. It must instead be the result of fallacious philosophical reasoning about what must be the case rather than everyday observations of what is the case.

#### No epistemology

Ryle's second objection is that if volitions really were internal mental acts, then it would never be possible to determine whether someone else's movements were caused in the right way to count as actions or not. The problem can be put this way. Imagine that you had no reason to believe whether the sounds and motions of those around you were deliberate actions or merely reflex movements. What evidence would be available to answer this question if the Myth were true? One could not look into another's mind to find out whether there were volitions causing their bodily movements. Even if they 'reported' that this were so it would not settle the matter. That would only help if one could determine whether the sounds made really were a sincere report that 'meant' what we would ordinarily take them to mean rather than being an exactly similar sounding but pattern of noises, which were not intended to convey any such thing. In other words, a report only helps if one already knows whether it is the result of a volition. But that is what is at issue. This objection is clearly a form of the Problem of Other Minds (see chapter 27).

#### No connection

The third objection is that the connection between volitions and actions is utterly mysterious. Ryle's objection here focuses specifically on a dualistic Cartesian picture that makes the interaction between the two sorts of substance particularly obscure. But it is worth noting as a more general challenge to monistic causal theories of mind. (Which is not to say that such theories cannot answer it.)

The mysteriousness of the connection between volitions and actions underpins the further epistemological claim that even in one's own case, one cannot know whether one's movements are actions. Even if introspection allows the detection of an inner volition preceding trigger pulling, that does not settle whether the volition really did cause the pulling. (We will see a related problem for causal theories later framed not in epistemological terms but as a problem in the analysis of action. If an action has

by definition to be caused by an appropriate mental event, what is the precise nature of this relation? As we will see, there are some cases where mental states do seem to cause movement but cause them in the wrong way to count as deliberate actions.)

#### *No answer*

Fourthly, Ryle raises a dilemma for the Myth of Volitions. If volitions are mental happenings, are they themselves deliberate or not? If they are, then by the theory they should themselves be the result of antecedent volitions and this generates an infinite regress. If they are not themselves voluntary, then it seems absurd that the 'actions' they cause are voluntary. If a volition were inserted into one's mind by an outside agency and that volition were sufficient to cause an action, would the action be deliberate? Ryle suggests not.

#### **Ryle's positive account**

Having spelt out his critical attack on one philosophical explanation, Ryle goes on to offer some positive remarks about the distinction. He focuses mainly on the distinction between those actions that are *blameworthy* and those that are not. What might otherwise be a morally culpable action can be excused if it could not have been helped in the circumstances, if it was, in *that sense* involuntary. But one does not ask whether admirable actions are voluntary, are anyone's fault. Furthermore, Ryle suggests, the issue of voluntariness in this context is satisfied without appeal to inner mental items. One discovers instead the standing abilities and capacities and details in which they were exercised. Clearly, this sort of idea would have consequences for thinking about agency and freedom in the context of mental health.

#### **Assessing Ryle's account**

There are two things to consider when assessing Ryle's positive account. One is whether he gives a satisfactory account of the distinction on which he focuses. Does the appeal to the broader, generally social, context of the use of the distinction between the voluntary and involuntary suffice for an explanation of that distinction? We will return to this issue in the final reading of the chapter (linked with Exercise 7), by Sir Peter Strawson (1974).

The other issue is whether the distinction on which Ryle focuses captures the pre-philosophical puzzle with which we began. Ryle suggests that philosophers have stretched the use of 'voluntary' and it is only because of this that they feel the need to postulate volitions. Effectively, Ryle suggests that a Wittgensteinian move is called for. The philosophical problem that the myth of volitions is supposed to solve should instead be *dissolved*. Philosophical *therapy* is needed rather than philosophical theory. We will discuss this a little more below.

#### **An issue may remain**

But despite the claim that philosophical puzzlement stems from distorting the genuine and substantial question of whether an action is blameworthy, there does seem to be a different and genuinely puzzling issue, which remains even if that area is clarified.

What remains is the distinction between actions and mere movements. This distinction is also mentioned in the reading (Ryle, 1963, pp. 71–72). It is the distinction between something that a person does and what is done to her. Ryle says: 'Sometimes the question "Voluntary or involuntary?" means "Did the person do it, or was it done to him?"; sometimes it presupposes that he did it, but means "Did he do it with or without heeding what he was doing?" or "Did he do it inadvertently, mechanically, or instinctively, etc?" (p. 72).

So although Ryle offers good arguments against the Myth of Volitions, it is less clear whether he provides a resolution to the philosophical question: What is the difference between actions and mere movements? Note that this question is not the same as: How do we know whether something is a voluntary action rather than an involuntary movement? Rather, what does that distinction consist in?

It also worth reflecting here on the extent to which Ryle's argument against the Myth of Volitions, which he construes as involving a causal element, succeeds against causal theories which do not rely on an immaterial ghost in the machine. Think of reasons as material states of the brain, capable of causal pushes and pulls and think of actions as movements, which are caused by such reasons. Do Ryle's four arguments undermine this picture?

It is hard to give a precise answer to this question. But a modern causalist might argue in the following way. There is evidence for this new version of volitions: they are the reasons for our actions which we often report in daily life and we distinguish between those reasons on which we act and those which are idle fancies. We can tell when others have such states by asking them because the effects of reasons are detectable in the way that many hidden causes are detectable in the physical sciences: by their effects. The connection between reasons and causes is relatively straightforward. Whatever complex physical states mental states really are, they cause, in whatever normal account we have of causation, actions. Finally, there is no dilemma. Our actions are voluntary if we act for our reasons. Now although we may rationally scrutinize our reasons, many of our mental states are not themselves only the result of our own past actions. Looking around may be a deliberate action, but what I see is independent of me, and on the basis of what I see I will form beliefs that lead to actions. But this does not threaten the voluntary status of my subsequent actions.

In fact things are not as simple as these replies suggest. But to get a deeper understanding of the debate, a short summary of the work of another non-causalist, criticized by the causalist Davidson, will be helpful.

#### **Melden's non-causal account of action**

##### **Melden's free action**

In his book *Free Action* (1961), in particular chapter 13, the philosopher A.I. Melden sketches a Wittgensteinian picture of action that eschews causation.

One theme is an important metaphilosophical and methodological claim. Melden suggests that a major source of philosophical error is a natural tendency to 'suppose that the difficulty we may have in understanding what an intention or desire is, is the difficulty involved in the *discovery* of an elusive item in our experience' (p. 172, italics added). The idea is that, with this strategy in place, the solution to philosophical problems is taken to require *inferring* the existence of private mental items to explain the observed phenomena: public actions.

### Action and meaning

Melden (1961) likens this approach to the nature of action to a similar move to explain meaning. There it involves the postulation of meanings or some other 'mental processes that ride piggy back, so to speak, on the words we utter' (p. 173). In that latter case, Wittgenstein's criticism (in his 1953 *Philosophical Investigations*) of mentalistic explanations of what needs to be added to sounds or symbols for them to possess, or for hearers to understand, their meaning has been increasingly influential. But Melden suggests that a similar moral applies also to the distinction between actions and mere movements. As in the case of special mental meanings so in the case of postulating internal volitions to explain actions: 'no such events could exhibit the requisite logical features of the concepts we employ' (p. 175).

Melden suggests instead (largely without further argument) that one should refrain from 'converting a question about meaning into an extremely questionable view about matters of psychological occurrence' (p. 175). Philosophers should instead 'examine carefully the manner in which terms like 'intention' and 'desire' operate in our familiar discourse about actions and agents' (p. 175). This approach should remind you of Austin's work, discussed in Part I.

### The conceptual-connection argument

With these methodological preliminaries in place, Melden advances the following key claim. Because there are *logical* connections between mental items such as motives, desires, and intentions and actions, *causal* theories of action must be false. The logical connection or dependence is manifested in two ways. First, it is impossible to understand the former concepts independently of understanding the latter and vice versa. Secondly, *explanation* of action by appeal to such mental states differs significantly from the causal explanation of events in the physical sciences. Melden fills out the second a little later. He says:

Where we are concerned with causal explanations, with events of which the happenings in question are effects in accordance with some law of causality, to that extent we are not concerned with human actions at all but, at best, with bodily movements or happenings; and where we are concerned with human action, there causal factors and causal laws in the sense in which, for example, these terms are employed in the biological sciences are wholly irrelevant to the understanding we seek. The reason is simple, namely, the radically different logical characteristics of these two bodies of discourse... (p. 184)

The underlying suggestion is this. The concepts of motive and action fit together with others in a structure that is different from and independent of the structure that relates cause, effect, and law. We have already encountered this idea in the distinction between the 'space of reasons' and the 'realm of law', which McDowell takes from the work of Sellars (see Part III). Melden argues that a causal theory of mind conflates two different conceptual structures. We will return to these arguments in the next session in order to assess both them and Davidson's counter arguments. But they are complemented by a further metaphilosophical methodological suggestion.

### The primacy of persons

Melden suggests that the philosophy of action specifically, and of mind more generally, should begin with, or presuppose, the concept of a person: a practical being with a primitive ability to move his or her limbs and to act for reasons. Furthermore, persons or people typically act within a *social* framework of rules and customs. Again this starting point is proposed without much argument. But it is worth contrasting it with Descartes' starting point. Descartes assumes only first person access to the content of one's mental states and to sensations. The starting point is a purely mental point of consciousness. As we will see (in Chapter 27) there are strong arguments against the intelligibility of first person access independent of third person access or of the concept of an ego independently of that of an embodied person. Melden subscribes to these Wittgensteinian views and concludes that a satisfactory philosophy of mind should reject Descartes' abstraction and the explicitly epistemological concern that goes with it. Starting points in philosophy by definition resist justification. But one could argue meta-philosophically that Descartes' starting point has been tried throughout 300 years without great success and that this is reason enough to start somewhere else.

Given his starting point, Melden claims that the route to a clear understanding of action is clear. It turns on an analysis of the concepts used to justify or explain actions in everyday life. One should examine the broader space of reasons deployed in justification rather than events that take place at the very same time as actions. The latter focus is perfectly acceptable but only as an explanation of bodily movements, perhaps as pieces of physiological or neurophysiological inquiry.

### Epistemology and interpretation

This leads to a further strand or argument. How is it possible to recognize an action as an action rather than as a mere movement? Melden likens this question to the question of how it is possible to see marks on paper or sounds as meaningful words. With that analogy in mind, Melden argues that it cannot be the case that one has to *interpret* movement as action. If that were so, he argues, there would be nothing left to justify an interpretation. Everything would hang in the air without conceivable support. This hint at an argument is a version of Wittgenstein's argument against using interpretation as the theoretical basis of meaning

discussed in Chapters 24 and 25. In this case, however, one may ask what would be wrong with a theory that related inputs described in meaning-free terms with meaningful interpretations of them. Is it obvious that this is incoherent? May be this is not the best explanation of our ability? We will postpone discussion of this question because it is a matter of the *epistemology* of mind rather than the ontology of mind and the nature of action explanation. It will be the subject of Chapter 27.

This session has outlined two influential arguments that action explanation is not causal. If those arguments hold good, then the connection between the freedom that is a key aspect of the space of reasons—the metaphorical space that characterizes action explanation among other things—and causal determinism remains mysterious. But in fact philosophical accounts of action explanation have moved on and now causal theories hold greater influence. It is to these that we will now turn.

### Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. What, according to Ryle, is the Myth of Volitions and what is its purpose?
2. What criticisms does Ryle offer of that myth and what positive account does he propose in its stead?
3. How does consideration of speech and meaning shed light on action according to Melden?

## Session 3 Arguments for a causal theory of mind

### The causal theory and reductionism

This session will outline the origins of recent causal theories of mind in Donald Davidson's seminal article 'Actions, reasons and causes' (1980 pp. 3–19). It is important first, however, to note a potential ambiguity in the use of the label 'causal theory of mind'. In some contexts it is used to refer to attempts to explain intentional concepts such as belief and meaning in non-intentional and usually causal terms. This is also called the project of 'naturalising' intentionality or providing a 'causal semantics' and was the subject of Chapter 24.

But even if one grants that that reductionist project is misguided, one can still subscribe to a causal theory of mind when this is construed in a different sense. One can argue that, although the intentional properties of mental states cannot be reduced to causal or other non-intentional properties, nevertheless the states or events that bare those properties still play a *causal* role in the

production of both action or other mental states, and are themselves caused by other mental states and perceptions.

Davidson's Anomalous Monism is perhaps the most influential example of such a theory. As we saw in Chapter 23, Davidson attempts to reconcile the view that mental concepts cannot be structured in natural laws with the view that the mental events to which they apply are still part of the causal fabric of the world. In what at first seems a counter-intuitive account, he attempts to have his cake and to eat it. We will see below that a key element of this causal theory of mind is that action explanation by citing reasons or propositional attitudes, is a species of causal explanation.

Although Chapter 23 introduced the metaphysical system by which Davidson hoped to reconcile the irreducibility of the mental with granting it a causal role, it did not discuss Davidson's argument for thinking that mental states are causes. That is the purpose of the next reading.

### EXERCISE 4

(30 minutes)

Read the extract from:

Davidson, D. (1980). Actions, reasons and causes. In *Essays on Actions and Events*. Oxford: Oxford University Press, pp. 3–19. (Extract: pp 11–17)

Link with Reading 26.3

Summarize (on a piece of paper) Davidson's counter-arguments to previous arguments against granting reasons a causal role.

- ♦ What positive considerations does Davidson advance for saying that reasons actually are causes (as opposed simply to conceding that there are no good arguments against them being)?
- ♦ Can you think of a different response to Davidson's challenge?

Davidson's paper advances two general theses about action explanation via reasons:

1. For us to understand how a reason of any kind rationalizes an action it is necessary and sufficient that we see, at least in essential outline, how to construct a primary reason.
2. The primary reason for an action is its cause.

The first half of the paper (sections I–III) discusses the first while the second (section IV) focuses on the second: on the causal role of mental items.

The key idea that Davidson wants to defend is that an intentional action is one done for a reason. But to know that an action was intentional is not in itself sufficient to know the full reason for the action. To know that the action is intentional is to know that there is some reason, but not which one it is. Actions can be undertaken with different aims in view, different ends or effects

of the same action. And even the same end can be desired, wished for or lusted for.

By giving a reason for an action the action can be revealed as coherent, rational, connected with broader patterns of action: When we ask why someone acted as he did, we want to be provided with an interpretation. His behaviour seems strange, alien, outré, pointless, out of character, disconnected; or perhaps we cannot even recognise an action in it. When we learn his reason, we have an interpretation, a new description of what he did which fits into a familiar pattern. (pp. 9–10)

### Intentions and intensions

One reason why knowledge that an action was intentional is not in itself sufficient to imply the reason for which it was carried out is that intentional actions are also intensional with an 's'. (What follows turns on one particular way of individuating or counting actions; see shortly below.) The very same action can be described in many different ways. The action or event of turning on the light may also be an action or event of warning a prowler. The same event can be described in these different ways. But although the action is intentional under one description, it may not be under others. I may want to flick the switch and to turn the light on, but I may have no thoughts at all about warning any prowlers. I do not intend to warn the prowler although that is what I do by turning on the light. Davidson's thought here is that an action is intentional providing that there is some description under which fits into a reason explanation.

(That way of putting it clearly turns on an assumption about how to count actions. Davidson suggests that we count actions as events that may have different descriptions depending on their relational properties. Because the action of turning on the light has the broader consequence of alerting the prowler, it can also be called that. More narrowly, it can be called the action of flicking the switch. Narrower still it can be called the action of moving one's finger. That the switch is flicked depends on the causal structure of the world. It turns on the world doing us a favour. (In fact Davidson elsewhere suggests that we pick out a subclass of 'basic actions' that turn on no favours and do not extend beyond the skin.) But Davidson's approach to action-individuation is to group these all together as different relational ways of specifying the same action. In much the same way, the same individual can be somebody's brother, somebody's son, somebody's father. But an alternative approach is to say that these are all different actions because they all exemplify different properties. Thus turning the key and starting the car are different actions. In what follows, action individuation will play no role. For a discussion see the introduction to A.R. Mele *The Philosophy of Action*, 1997.)

### Reasons as rationalizations

So far, Davidson's (1980) account of reason explanation sounds much like the broadly Wittgensteinian account suggested by both Ryle (1963) and Melden (1961). It presents action explanation as a method of fitting one action into a broader pattern. Reasons are

an appropriate spoonful of contextualizing information. They provide the resources for redescribing the action to be explained. Davidson formalizes the rationalizing property of reasons thus: 'C1. *R* is a primary reason why an agent performed the action *A* under the description *d* only if *R* consists of a pro attitude of the agent towards actions with a certain property, and a belief of the agent that *A*, under the description *d*, has that property.' (p. 5).

But he goes on to add to that picture an explicitly anti-Wittgensteinian element. He claims that reason explanation is also a form of causal explanation. The reason for an action is (part of) its cause. In order to ground this claim he needs to do two things. One is to overcome the (largely Wittgensteinian) arguments previously deployed to show that reasons cannot be causes. And he also needs to provide positive grounds for thinking that reasons actually are causes.

### Against non-causalism

With respect to the first target, Davidson provides two key counter-arguments:

1. *Non-causalist claim.* The normative connection between mental states and behaviour is an obstacle to any causal account. Reasons and actions have to be described in a suitable way in order to display their rationalizing powers. This implies that there is an analytic or logical connection between reason and action whereas a causal link is contingent and empirical.  
*Davidson's response.* The analytic connection depends on how the facts are *described*. But the causal relation does not. The fact that an analytic connection can be made by a suitable choice of description cannot preclude a causal connection because any given causal relation can be so described. Suppose event *A* causes event *B*. Event *A* could be described as 'the cause of *B*' in the analytically true statement: the cause of *B* caused *B*. But this does not contradict the assumption that *A* caused *B*.
2. *Non-causalist claim.* The statement that someone acted in a particular way because of a particular reason does not imply that reasons of that type *generally* lead to actions of that type. But causation requires just such generality.

*Davidson's response.* A Humean or nomological account of causation requires that some description of the two events is possible that connects them as a matter of law. But it does not require that the descriptions used to pick out cause and effect are suitable for inclusion in the law. It may be that the cause of an event that is reported on page 13 of the *Tribune* is itself reported on page 5 of *The Times*. But that does not imply that there is a linking law that uses the descriptions: 'events reported on page 13 of the *Tribune*' and 'events reported on page 5 of *The Times*'. Ignorance of the actual law does not, however, eliminate causal explanation.

Both of these counter-arguments are based on the following strong intuition or assumption. Unlike the intentionality of

actions, their causality is extensional rather than intensional (again with an *s*). If one event causes another it does not matter how it is described. The underlying metaphysical facts about the causal structure of the world are indifferent to how we describe them. By contrast one may *not* have intentionally or deliberately killed one's father even if one deliberately struck an intruder over the head under circumstances where as a matter of fact in striking one killed and the intruder was one's father. Whether the action was intentional depends on how it is described. (Again, this account follows Davidson's way of individuating actions.)

Once the claim that causality is an extensional or description-independent relation is in place, Davidson has a general method for defusing the kind of argument that Melden deploys. Although the conceptual structure into which 'reason' and 'action' fits—the space of reasons—differs from the conceptual structure corresponding to the realm of law and including 'cause' 'effect' and 'law of nature' this does not imply that the two structures cannot apply to the very same items. They may structure them in different ways but, Davidson suggests, what is being structured is the same in the two cases.

As we have already seen (in Chapter 23), Davidson goes on to suggest that the (mental) events labelled by the mentalistic vocabulary of propositional attitudes just are the very same events that can also be picked out using the vocabulary of causes, effects, and strict or probabilistic laws. *Really*, there are only events and these can be picked both via their mental and also by their physical relations or properties. It is just that these two systems cannot be systematically related. But before accepting the need for his unifying metaphysics of events, we should ask a further question. What positive reason is there to think that reasons are causes?

### Reasons as causes

Davidson provides two positive arguments. We will pick out one here and discuss the second below.

The first takes the form of a challenge to non-causal theories of mind based on the question: what is the difference between a reason for an action and *the* reason? The challenge Davidson puts forward is to explicate the difference between a case in which someone has a reason for an action *and* carries out that action—where the 'and' is read purely conjunctively—from cases where they act *because* of the reason, where the reason is 'active': '[A] person can have a reason for an action, and perform the action, and yet this reason not be the reason why he did it. Central to the relation between a reason and an action it explains is the idea that the agent performed the action *because* he had the reason.' (p. 9).

The philosophical issue here is to explain what this distinction amounts to. What constitutes the reason for an action as *the* reason?

This problem is not a matter of epistemology. It is not a question of how one *knows* which reason is *the* reason for an action. Davidson's implicit assumption is that epistemology is little guide to ontology here. How in practice one knows what the reason for

an action is, either in one's own case or for others, plays no part in the discussion. A philosophical account of the distinction need not postulate facts which help *guide* practical knowledge. (On the other hand, the account should not make knowledge here utterly mysterious or impossible.) Nor is the objection that one may always act for a number of reasons relevant. Providing that it is possible to act, and to have even one reason for that action which is *inactive*, then an account is owed as to what the difference between an active and an inactive reason is.

### A reason versus *the* reason

Davidson argues that the rationalizing aspect of reasons is no help here. As we have seen, the rationalizing role of reasons is to make sense of action through contextualization. But if the rational power of reasons is understood in the way Davidson suggests and formalizes as condition C1 (above), it cannot constitute the difference between *a* reason and *the* reason. To be *a* reason is already to be a reason that is held by the person who acts and which rationalizes the action. Rationality has already been used in distinguishing between such reasons for action and mental states that have no bearing on the action whatsoever. It cannot also provide the extra ingredient sufficient for being the actual reason why someone acted.

Davidson suggests that one might augment the characterization of rational power so as simply to include the extra ingredient, whatever it is. But the cost of that assumption would be to make the rationalizing role of reasons mysterious. Davidson suggests that the rationalizing role should instead be construed in the transparent way already described and that something has to be added to the rationalizing role of reasons to answer the question.

### Causation as the extra

Davidson's proposal is that the extra ingredient is causal efficacy. *The* reason for an action is the reason that causes it. Thus reasons have to meet a second non-rational condition: 'C2. A primary reason for an action is its cause.' (p. 12).

Thus the argument that Davidson puts forward is of the following form. Something needs to be added to the rationalizing force of reasons to distinguish between *a* reason for an action and *the* reason. The only candidate for this extra ingredient is causation. One way of putting this is that Davidson points out that there is a second necessary condition on reason explanation. But he gives no reason to believe that it is more than contingently true that this condition is satisfied by causation and little specific reason to believe that it is true of causation.

There is, however, implicit in some of his remarks in this paper a further argument why causality specifically has to be added to rationality to account for reason explanation. This has recently been made more explicit in a commentary by William Child (1994).

### Child's version of the argument

In his book, *Causality Interpretation and the Mind*, Child (1994) develops and defends a causal theory of mind as that phrase is

being used in this chapter: mental states play causal roles without any implication that mental properties can be analysed in causal terms.

### Causes and explanation

Child reiterates the Davidsonian argument already described, but also goes on to make explicit another argument that is only hinted at by Davidson. This argument can be summarized baldly as follows:

1. an action explanation is an explanation of why something happened; but
2. no non-causal explanation can explain why something happened; so
3. action explanations must be causal.

Clearly this argument turns on the plausibility of its second premiss. There are two ways of testing this. One is to try to devise explanations that do not turn on citing causal explanations of why *non-mental* events occur. The other is to question whether the explanation of mental events and that of non-mental events must have the same underlying logical form.

With respect to the former, Child (1994) makes the following point:

The idea that any explanation of the occurrence of an event must be causal is plausible case by case. For any putatively non-causal explanation of an event, we can always make the Davidsonian point: knowing this story allows us to fit the event into a pattern which potentially makes sense of it; but we are still left wanting to know why the event actually occurred, what made it happen when it did. (p. 92)

Thus Child proposes the following challenge to anyone who disputes the second premiss. Find a non-causal explanation of an event that is not merely an explanation of what constitutes it as the sort of event it is and really is an explanation of why it occurs. His claim is that all the non-causal explanations alluded to by the 1960s Wittgensteinians, for example, tell us something about the events in question but not why they happened.

Of course, this sort of argument has an essential modesty. Because Child cannot investigate every putative case of non-causal explanation, and because the only general rule he offers is the very one which is in dispute, he cannot provide a watertight justification of this principle. Nevertheless, his argument does provide strong intuitive support because of the *de facto* success of his challenge. For non-mental events, explanations of why they occur do seem to require the specification of causal information.

### Child and Lewis

Child's argument is in accord with Lewis's account of natural scientific explanation, which was discussed in chapter 14. Lewis claims that all such explanation relies on the provision of part of the causal history of events. One of the criticisms made of Lewis's account there was that it was incomplete and did not cover all forms of explanation, such as mathematical or logical explanations

which cite proofs. Child's argument is more limited because he restricts his concern to a subset of explanations: those concerned with why events occur, rather than anything else about them, and thus he escapes that sort of criticism.

But there is a second kind of response available. This is to concede Child's general claim with one important class of exceptions: the class of mental explanations. In other words, anyone who doubts that action explanation by reasons is a species of causal explanation might concede what Child claims *outside* that sphere but continue to dispute the relevance of those cases to the mental case. Even if explanations of why Atherton was bowled Leg Before Wicket, which explain why that event happened—by citing how the ball bounced on the dry earth, its speed, and the length of human reaction times—that need not show that *reason* explanations are causal. The explanation of why the bowler decided to try that kind of bowl by appeal to his plans, perceptions, and desires need not be causal.

To assess the force of this response requires further reflection on the nature of the claim that reasons are causes or that causality provides for a distinction between action and mere movement. As the next stage in this process of clarification, the following session considers the kind of causal theory of mind involved. Davidson argues that actions are caused by mental events while others argue that there is instead just a brute relation between an agent and their action.

### Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. What are the two general ingredients that Davidson thinks are involved in action explanation (hint: not in this context belief and desire!).
2. What specific argument does he offer for the second element?
3. How does he defuse Wittgensteinian arguments against this?

### Session 4 Event causation, agent causation, and irrationality

It is one thing to argue that reason explanation must be a species of causal explanation and another, further claim to explain how precisely this comes about: what reasons are or what is causally related. This session will look at two contrasting versions of a causal account. One is Davidson's account, already discussed in Chapter 23 in which causation links free-standing mental states and actions. The other is an account in which causation links a whole person to his or her actions.

### Recap of Anomalous Monism

As we saw in Chapter 23, Davidson provides a novel reconciliation of three initially apparently irreconcilable claims:

1. that mental states stand in causal relations;
2. an underlying nomological account of causation;
3. the anomalism of mental states.

These appear to be in tension because the first and second together imply that there are laws that govern mental states while the third claims that there are no *mental laws*: no laws (or law statements to be precise) couched in mental terms. Davidson's solution is to claim that the very same states that are picked out using the mental vocabulary of reason explanation, can also be picked out in principle by their physical properties, which can be fitted into a structure of strict physical laws. The third claim remains true because there is no systematic way of relating mental properties to physical properties in such a way that the mental might be reduced to the physical. If there were bridge laws connecting mental and physical properties then given the existence of strict physical laws, there would also be strict mental laws. Thus to preserve the truth of the third claim—that there are no such laws—Davidson is forced to claim that there are no psychophysical laws.

Chapter 23 charted some of the objections that have been raised to Davidson's account of the metaphysics of mind. These include the objection that Anomalous Monism cannot provide an account of the causal efficacy of mental *properties* to which we will return shortly. But, on the initial assumption that these objections could be overcome, one strength of Davidson's account should now be clear. Davidson's general claim that reasons are causes relies on a straightforward construal of what causes are. Davidson simply means the same kind of causal relations that connect other non-mental events. Furthermore, he appeals to a familiar broadly Humean nomological account of causation. Thus the attraction of Davidson's account is the attraction of unification. The causal relations between mental states or between mental states and actions are of just the same kind as those described in the rest of the physical sciences. In saying that causation is an element of an account of agency, he means just the sort of causal relation that can connect events generally.

### Agent-causation

Davidson's is not the only model of a causal theory of mind. Child (1994), discussed above argues for a causal theory but does not think it should link mental states to actions. Similarly, John Bishop Professor of Philosophy at Auckland defends a very different kind of causal theory of mind in this paper 'Agent-causation' (1983). On his account of agency, the causal relation connects not mental events or states to other states and actions, but, more primitively, it connects the agent herself to something else. (What is the other relatum? As he explains, the agent-causal relation cannot connect the agent to her action because the action is inclusively defined as the holding of the relation. He suggests that it connects her to her movement, narrowly described. The action

is thus the whole fact of an agent standing in an agent-causal relation to a movement.)

There is a clear disadvantage of this approach as the contrast with Davidson should highlight. It does not provide the kind of unification to which Davidson's nomological event-causation account aspires. This follows from the fact that agent-causal relations—whatever they are—are not covered by standard nomological accounts of causation. Bishop defends the approach by arguing that the element of reduction implicit in Davidson's project fails because of an anomaly with which it cannot cope. Agent-causation is embraced as an alternative without any such reductionist intent. (Note that talk of reductionism here does not imply the reduction of mental-content to non-intentional concepts as it did in Chapter 24. It refers to the reduction of the concept of action to more primitive concepts such as rationalization plus causation.)

### Causal deviance

The anomaly on which Bishop builds his argument is sketched out with admirably honesty by Davidson himself in a different essay:

Let a single example serve. A climber might want to rid himself of the weight and danger of holding another man on a rope, and he might know that by loosening his hold on the rope he could rid himself of the weight and danger. This belief and want might so unnerve him to loosen his hold, and yet it might be the case that he never *chose* to loosen his hold, nor did he do it intentionally. It will not help, I think, to add that the belief and the want must combine to cause him to want to loosen his hold, for there will remain the *two* questions *how* the belief and the want caused the second want, and *how* wanting to loosen his hold caused him to loosen his hold.

Davidson 'Freedom to Act' in *Essays on Actions and Events* (1980, p. 79)

Here is the problem. The first of the two arguments for a causal theory of mind isolated above took the form of a challenge. What has to be added to mere rationalizing force to distinguish a reason for action (which the agent may have) from *the* reason why she did the action? Davidson's suggestion is that causation is the relevant extra. But while causation may be a *necessary* extra ingredient for action, the example of the climber shows that it is not a *sufficient* addition—when added to the possession of reasons. In this example, and a host of others, the possession of reasons causally leads to some behaviour but leads to it in the wrong sort of way for it to count as intentional. The philosophical problem is that of spelling out what the right sort of way is which does not simply presuppose the concept of action.

To the normal range of cases that philosophers consider, we can add psychopathological cases such as made actions, made impulses, and so on. In such cases, subjects can 'act' and can have reasons for those actions which, if a causal theory is right, cause them. But there is still something pathological if the subject denies that they are the authors of their reasons.

In 'Agent-causation' (1983) John Bishop argues that the specification of causal deviance will always rely on a more primitive account of agent-causality. Thus in the climber example: 'He didn't let go

intentionally, even though his intention to let go caused him to lose hold, because his nervousness prevented him from controlling his grip' is a reasonable off the cuff account of the deviance... [which] makes appeal to a notion of maintaining control which may well have agent-causal presuppositions.' (p. 68).

In general, diagnosis of which event-causal chains correspond to failures of intentional action will always rely on a prior and primitive theory of agent-causation.

### Agent-causation and deviance

The substance of such an agent-causal account of agency turns on the nature of the primitive theory of agent-causality. Bishop (1983) spells this out as follows:

[F]or M to Q intentionally, where this is a basic action, M must have the intention to Q, and M must exhibit behaviour which counts as Q-ing. But we require furthermore that, in behaving as he does M should himself be carrying out his intention to Q... Its satisfaction itself requires two further conditions be met, the *real capacity condition*, and the *non-pre-emption condition*...

There is a natural explication of such pre-emption... If, and only if, there are antecedent conditions causally sufficient for M to exhibit Q-ing behaviour which are *independent* of M's carrying out any proximate intention he may have to Q, then, if M *does* form such an intention, his opportunity to carry it out is pre-empted. (pp. 75–76)

Bishop explicitly concedes that this sort of primitive theory is neither meant as a *reductive* explanation of intentional action nor deviant causation. Nevertheless, it is appropriate to ask what sort of elucidation Bishop's sketch of a theory succeeds in providing. The worry is this: it relies upon an unexplained notion of sufficient conditions for behaviour *independent* of intentional action. This casts doubt on the claim that agent-causality provides any unpacking of the notion of agency itself. While Davidson argues that light can be shed on what makes an action intentional by appeal to its causal origins, Bishop relies on a prior notion of intentional actions in order to diagnose whether behaviour caused by antecedent mental states is intentional or not.

### But why agent-causation?

On the assumption that some form of causal theory of mind must be true and that causal theories must be of one form or another then Bishop's argument against Davidson's event-causality provides an argument for agent-causality instead. But this leaves another possibility. One might choose instead to reject the assumption that causality has anything to do with agency. As on Bishop's account causality only ever features as part of the primitive composite notion of agent-causality, why think of this as any form of causality rather than, say, agency? The primitive theory that he offers to form the backdrop for diagnosing what is wrong in the climber example could remain but reconstructed as a piece of conceptual analysis of what we all implicitly understand by agency. The next session will return to this possibility. For now we will consider a further difficulty of Davidson's causal theory. This

turns on the distinction between mental content and the vehicle of mental content set out in Chapter 24.

### Davidson on irrationality

The following extract (linked with Exercise 5) sets out a framework which Davidson (1982) suggests might help make sense of certain forms of irrationality: centrally weakness of will or 'akrasia'. It relies on the fact that Davidson's account allows a distinction between the rational and causal properties of mental states. One of the themes of this paper is that this distinction has consequences for the intelligibility of akrasia. A second is that any plausible account will also require the partitioning of the mind. For now, the first will provide a point of criticism of Davidson's account that shows a principled difficulty for causal theories of mind.

#### EXERCISE 5

(30 minutes)

Read the short extract from:

Davidson, D. (1982). Paradoxes of irrationality. In *Philosophical Essays on Freud* (ed. J. Hopkins and R. Wollheim). Cambridge: Cambridge University Press, pp. 289–305. (Extract: pp. 296–298)

Link with Reading 26.4

- ♦ What is the philosophical difficulty with understanding the very idea of irrational actions or beliefs?
- ♦ What is Davidson's key idea for accounting for them?
- ♦ What does this suggest about rational beliefs and actions?

### Akrasia and paradox

As Davidson points out, the paradox of irrational actions or beliefs is that they are failures *within* what McDowell calls the space of reasons. If instead they were simply *non-rational* they would lie outside the sphere of rationality completely and would not be paradoxical. But irrational acting or thinking is subject to reason explanation and thus subject to the inbuilt rationality that that form of explanation carries. Irrational actions are, however, subject to merely *partial* reason explanations: reason explanations that fail to be fully rational. The philosophical difficulty is to account for this halfway house.

Davidson's solution is that irrationality is the result of reasons whose causal efficacy pathologically exceeds their rationalizing force. Taking the case of a man who goes out of his way to replace a branch he had moved in a park:

The man who returns to the park to replace the branch has a reason: to remove a danger. But in doing this he ignores his principle of acting on what he thinks best, all things considered. And there is no denying that he has a motive for ignoring his principle, namely that he wants, perhaps very strongly, to return the branch to its original position... Irrationality entered when his desire to return made him ignore or override his principle. (p. 297)

Whereas in normal cases, beliefs stand in both rational and causal relations to each other and to actions, in cases of irrationality some of the rational relations are distorted or overridden by merely causal relations. In the case of wishful thinking, for example, a reason—the wish to have a particular belief—which fails to offer rational support for that belief, serves as a reason, and is causally sufficient, for holding that belief. Irrational behaviour is sufficiently intelligible for interpretation via rational reasons still to be possible. The relevant causes of behaviour are still mental states—reasons—with semantic content, as displayed by their other rational connections. But the causal power of some of these reasons exceeds their rational power.

This looks at first a promising strategy for accounting more generally for content-laden psychopathological symptoms. The problem for philosophical elucidation is to account for the fact that symptoms can be meaningful, causal, and pathological. It seems at first plausible that these three constraints could be met by a suitable development of Davidson's ideas. Symptoms can be interpreted as meaningful by an analyst who charts a network of rational connections between them and the sufferer's mental economy. But not all the connections are rational or fully rational. In some cases the causal powers of mental states are out of proportion with their rational powers. It is this that leads to some outputs counting as symptoms rather than rational actions or utterances.

### The mysterious harmony of causal and rational power

But a problem with this thought highlights a general problem with Davidson's reconciliation of reasons and causes. He provides no answer of why it is generally the case that the rational power and the causal power of reasons stand in proportion. There are simply no resources in his account to answer this question. As discussed in Chapter 23, Davidson fails to unite reasons and causes at the level of mental *properties*, which in turn casts doubt on the reconciliation achieved at the level of mental particulars.

This problem needs careful statement. As causal relations are extensional, it is not *in virtue of* its description as physical (or mental) that a mental event causes another event. Causal relations hold, or fail to hold, however the relata are described. Nevertheless, the properties that are invoked in the nomological account of the causal efficacy of mental events are exclusively physical. There are, according to Davidson, no strict laws linking mental properties. Thus mental *properties* play no part in causal explanations of action. But our standard model of the explanation of the occurrence of an event is that the properties that are cited—perhaps following the word 'because'—are causally relevant. It is in virtue of possessing those properties that the event happened. This is not true of Davidson's explanation of why reasons are causes. Thus he fails to unite reason explanation and causal explanation. The fact that events can play a role in both spaces is not sufficient to unite reasons and causes. While his account manages to display the rational structuring of reason and give an account of the causal role of reasons, it cannot explain how reason can itself play a role in causal explanation. He fails to reconcile or unite the rational and the causal.

### Causal efficacy and mental *properties*

This is another way of bringing attention to the issue discussed in Chapter 23. Can a merely token identity of mental and physical events or states accommodate a suitable causal role for mental properties? (This is the central subject of the edited collection J. Heil and A. Mele *Mental Causation*, 1993.) Of course it is no problem for an account that eschews a causal role for the mental (to which we will turn in the next session). Nor is it a problem for causal theories of the mind, which are also reductionist about mental properties. If mental properties can be reduced to non-mental physical properties using bridge laws then they will play a causal role of just the same sort as physical properties. But it is a problem for any account that aims to be non-reductionist about mental content while advocating a causal interpretation of action explanation.

One way of making this point reflects back on a distinction made in Chapter 24. It is the distinction between (mental) content and the vehicles of that content. In Chapter 25 the McDowellian objection was raised that any attempt to reduce mental content through talk of the properties of free-standing internal states or events, which encoded mental content was open to the objection that it left the mind dark. If such states stand in a network of 'horizontal' causal relations, which fully explains their occurrence, why also attribute to them 'vertical' relations of meaning and reference? Why assume that these go hand in hand? A related objection can be raised in the current context. Davidson attributes both causal and rational properties to mental states construed as internal free-standing events. But why assume that these go hand in hand? And why think of the causal properties as having anything to do with the rational properties of the content that the events or states encode?

Given the objections raised against reductionist accounts of mental content in Chapters 24 and 25 and given the objections raised against Davidson's weaker token identity theory embedded in Anomalous Monism, it is time to reconsider what arguments there are for a causal theory of mind (in the modest sense of this chapter: the claim that reasons are causes). In Session 5 we will look at an article that rejects Davidson's claim that there is a need to invoke causation to explain action.

### Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. Can the intuition that causation has something to do with agency be interpreted in ways other than Davidson's?
2. If so is it still clear that causation is playing a specific and characteristic role?
3. Can causation be used to explain irrationality?

## Session 5 A non-causal account of agency?

In the light of the internal difficulties of causal theories, this session will re-examine the arguments that have been deployed in favour of the claim that reasons are causes.

Just as Davidson's paper 'Actions, reasons and causes' (1980 pp. 3–19) undermined an assumption that was held largely independently of particular compelling arguments that reasons could not be causes, more recent work, following Davidson, has almost universally held the converse. While Davidson himself did not put forward genuine arguments for the claim, most philosophers of mind hold it instead as an article of faith, believing that Davidson's arguments are secure. The dominant approaches in philosophical accounts of mind such as functionalism or Representationalism, described in Chapters 23 and 24, are examples of this. Thus arguments directed against Davidson in a recent paper by Julia Tanney are especially interesting because of the consequences they would have if successful.

### EXERCISE 6

(30 minutes)

Read the extracts from:

Tanney, J. (1995). Why reasons may not be causes. *Mind and Language*, 10: 105–128. (Extracts pp. 108–110, 111–112, 113)

Link with Reading 26.5

Assess the strength of Tanney's counter arguments in the first half of the paper.

- ◆ What is the role that explanation plays in her choreography of Davidson in the middle section?
- ◆ What kind of account of mental explanation is suggested by her later remarks?

### Tanney's arguments against Davidson's arguments for reasons being causes

This paper begins by recapping the general motives for Davidson's construal of reasons as causes as outlined earlier in this chapter. Something has to be added to the fact that a reason rationalizes an action or shows how it is reasonable to distinguish *a* reason for an action from *the* reason. And what has to be added is causation because only it promises to give an account of the 'mysterious connection' between reasons and actions. Tanney then goes on to question whether causation does provide a satisfactory addition and whether any addition is in fact needed.

First, taking the case of Oedipus unknowingly killing his father, there is no need to add causation into the account because it is not needed to distinguish between his reasons for killing the threatening old man and any more 'Oedipal' desires to kill his father. The latter are irrelevant to explaining his actions because

he does not believe the old man is his father. Thus it does not serve as part of a reason *he* has for killing the old man.

In the second case of overridden reasons, Tanney assumes that Oedipus does know that the old man is his father and that in addition to his Oedipal desires (to kill him) he also has moral qualms about killing his father. Nevertheless, these qualms are overridden by his own desire to survive when threatened and so he kills his father. Given this scenario, Oedipus's Oedipal desires do not form part of his reason for killing his father even though they are reasons he has. They are not his reason because they are *overridden* by his moral qualms. It is only because, in addition, he has a desire to survive and believes that he is threatened that he kills his father.

Davidson would explain this by saying that the Oedipal reason is not causally active in this case: that causation makes the difference between a reason and the reason. But as Tanney points out this assumption is unwarranted once an account of competing reasons is given that trades only on rational and motivational concepts. A difference in the space of reasons can make the difference (between *a* reason one has and *the* reason one acts) instead.

A distinct motive for introducing causal powers in addition to rational powers might stem from thinking about Buridan's Ass who has equal reasons for choosing either of a pair of exclusive choices. If we want to explain why one is chosen rather than the other, we may want to say that the reason for it was causally active while the other was not. But Tanney argues that the right response to such circumstances is to recognize that there simply is no reason for a choice of one over the other and the choice is thus not subject to reason explanation. It is rational to *pick* one or the other rather than becoming fixated on this one choice, but picking in this case is not a rational choice but like an arbitrary toss of a coin.

Nor, finally, does the case of weakness of will justify introducing causal powers. In such cases there may be a causal explanation of sorts (hormone imbalance, perhaps), but this is not part of a reason explanation, in the sense of a reason an agent has for an action. It may be necessary to complicate a purely rationalizing account by introducing competing subsets of beliefs, but this is not itself causal.

### The role of weighted reasons

Tanney concludes that, once a more complex story of *weighted* reasons for actions is in place, there is no need to add a causal element to reason explanation. She then goes on, in Section 3, to suggest a deeper underlying motive for thinking that there must be some such causal addition. (She considers it only to reject it.) This is that there should be a determinate relation between reasons for action and action: that the former should be a sufficient condition for the latter. Tanney herself argues, on the basis of consideration of the Buridan's Ass case and that of weakness of will that this is too strong a requirement to place on rational explanation. Sometimes reasons are insufficient for action.

### A further problem

Section 3 of the paper continues with a further criticism of Davidson's attempt to combine a causal theory of mind with the

anomalousness of the mental. The first argument turns on the model of causation implicit in Davidson's account: Hempel's covering law account. Its basic idea is that whenever one event causes another there is a law-statement that taken with a suitable description of the occurrence of the first event implies that the second event occurs. Combining this with Anomalous Monism produces the following *prima facie* problem. As the laws that underpin the causality of the mental are physical, but the descriptions of events that feature in action explanation are mental, they will not 'fit together' to yield the right implications.

Now Davidson might respond to this objection in the following way. It does not matter that mental descriptions of events do not plug into the covering law account of causation. All that matters is that in principle there are pairs of suitable physical descriptions for the same events that do so fit. The cost of this response, however, is to sever the connection between this underlying metaphysical requirement and the reason explanations we give. As Tanney says of a slightly different point: 'But what any of this has to do with our original reason-attribution is left utterly mysterious since we haven't any idea how to identify the original event... as one that is apt both for the *appropriate* mental and physical predicates/properties to begin with.' (p. 119).

#### Where does this leave us?

So it seems both that Davidson's original arguments for a causal ingredient in reason explanation fail and that there would be something mysterious about the explanatory role of such an appeal even if further arguments could be found. Where does that leave us?

One possible line of response is to return to the sort of non-causal conception of mind with which the chapter started. Melden (1961) and Ryle (1963) argued that it was a mistake to think of mental states as the sorts of things that might stand in causal relations. Recall that Melden suggested that a major source of philosophical error was a natural tendency to 'suppose that the difficulty we may have in understanding what an intention or desire is, is the difficulty involved in the *discovery* of an elusive item in our experience' (p. 172, italics added).

But one virtue of Davidson's general account in the light of the first reading (linked with Exercise 1) by Spence (1996a) on neurophysiology was that it suggested a way of accommodating two different perspectives on action: the causal and the rational. It promised a picture of *how* freedom and determinism might be compatible by suggesting that these notions attached to different patterns of explanation of the very same events. In virtue of standing in a network of physical laws, mental events could be seen as physically determined. In virtue of standing in a rational pattern, they could be seen as expressions of autonomy. Furthermore, what makes mere movement into an action turns on its having the right kind of causal history: being caused (in the right way) by antecedent events. So one cost of giving up Davidson's causal theory of action and his token identity theory

of mental and physical events (i.e. there is just one set of events) is the loss of this response to the sort of worry raised by Spence (1996a). Recall, again, that this response was to suggest that Spence's challenge was just a specific instance of the more general problem of locating mind in nature and that Davidson provided that general account.

On the other hand, however, this kind of general picture of the dovetailing of the mental and the physical, with its core assumption about a causal theory of agency, might still seem to leave the general problem of reconciling freedom, which is a prerequisite of autonomous action, with the causal determinism highlighted by Spence. So in the last session we will return to Spence's paper and an influential paper by the Oxford philosopher Sir Peter Strawson.

### Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. Does Davidson really establish a causal theory of mind?
2. Can his arguments be overturned?
3. If so what remains of his account?

## Session 6 Freedom and determinacy

### The position so far

Let us summarize the position so far. Spence (1996a) highlights a seeming conflict between our ideas of conscious freedom and neurological determinism. Rather than tackling that question head on, we turned to a debate in the philosophy of mind about the nature of action, as that is usually taken to require the relevant sense of autonomy and freedom. Here there has been debate between the position that formed the orthodoxy in the UK and the more recent Davidson-inspired orthodoxy. The first position holds that there is no connection between the characterization of a movement as an action and its causal antecedents. The second holds that, to the contrary, it is in virtue of being appropriately *caused* by mental states, that a movement is an action. The rare occasions that the rational and the causal properties of a state diverge comprise examples of weakness of will and thus Davidson's picture suggests an explanation of that phenomenon.

But, as we have seen, there are problems with Davidson's attempt to keep the rational and the causal properties of mental states or events in harmony. What explains the rarity of weakness of will? This suggests that there is something wrong with Davidson's way of explaining the nature of action.

Furthermore, even given Davidson's picture of the mind, there may remain the suspicion that Spence is right and that freedom

and causal determinism are incompatible. This suspicion could be raised in the context of Anomalous Monism by saying that just because the rational and the causal properties of mental states are relatively autonomous, does not prevent causal determinism trumping the requirements of freedom at the rational level.

This final session will return to the question of freedom raised by Spence by examining an argument set out by the Oxford philosopher Sir Peter Strawson. This paper argues that, in so far as determinism is a coherent metaphysical doctrine, it cannot and should not threaten the idea of moral responsibility. Strawson argues for this without making any assumptions about the connection between mental and physical descriptions. In other words, if he is right, his conclusions apply independently of the correct view to take on whether reasons are causes.

#### EXERCISE 7

Read the extract from:

Strawson, P.F. (1974). Freedom and resentment. In *Freedom and Resentment, and Other Essays*. London: Methuen, pp. 1–25. (Extracts: pp 6–9, 10–11)

Link with Reading 26.6

- ◆ How does Strawson characterize the debate?
- ◆ What sort of response does he think should be given?
- ◆ Is he successful?

#### Strawson's twin aims

In this paper, Strawson attempts to carry out two related but separable tasks. One is to block the threat that determinism seems to pose to the idea of freedom and the various moral judgements that presuppose it. The other is to diagnose why it is that the choice between 'optimists' and 'pessimists' in that debate seems so stark. In this, Strawson aims at the therapeutic dissolution of a problem rather than providing a substantial justificatory philosophical theory to answer a genuine question. He wants to show why there is not a substantial problem concerning freedom, but without showing that we are in a strange sense *really* free.

#### Optimists and pessimists

Strawson characterizes the debate he wishes to assess as one between optimists and pessimists. Pessimists hold that if determinism is true then the concepts of moral obligation and responsibility lose their point. Optimists hold that this is not so. In fact Strawson further characterizes optimists as typically holding that moral concepts take their point from those moral practices that successfully regulate society. Both agree on the following further claim: that moral responsibility requires freedom. However, they disagree on what this freedom amounts to. Optimists think it is an absence of external compulsion and the presence of a coincidence of action

with the real wishes and intentions of agents. However, pessimists think that there is an extra sense to freedom that is incompatible with the truth of determinism, although, as Strawson points out, it is usually not possible to characterize this sense of freedom at all clearly. At this point an impasse is reached.

Strawson's therapeutic aim is then to provide the optimist with a further characterization of the grounding of moral concepts that will preclude the 'panicky metaphysics' of the pessimist. If something more can be said about the grounding of our moral concepts than merely their efficacy in regulating social practice, then perhaps there will be no need to invoke a queer sense of freedom to justify them.

#### Moral reactions

The first stage of Strawson's campaign (in section III of the paper) is to move away from more or less *intellectual* moral judgements to direct moral *reactions*. Here he notes that one's reactions to the actions of other depends greatly on their attitudes and intentions towards us. These affect whether one feels gratitude, resentment, and forgiveness, for example.

There are, however, occasions when one does not feel resentment, say, for an action that in other circumstances would prompt it. In one range of circumstances one takes the agent to be a morally responsible individual but the specific action to be one for which he or she was not responsible. Perhaps it was carried out under compulsion. In a second broad range of circumstances, one takes the agent not to be responsible: either because 'he was not himself' or because 'he's just a child' or suffering from schizophrenia. In this second range, of cases, instead of sharing a human relationship, one takes an objective attitude to the other: one treats them as someone to be managed, handled, or trained.

Strawson argues that we can also take this objective attitude to normal people, but not for long.

#### Moral reactions and determinism

With this context in place, Strawson goes on to ask what effect the truth of determinism should have upon our moral reactions or reactive attitudes. 'More specifically, would, or should, the acceptance of the truth of the thesis lead to the decay or the repudiation of all such attitudes? Would or should it mean the end of gratitude, resentment, and forgiveness?' (p. 10). This question amounts to whether the truth of determinism would or should justify the universal adoption of the objective attitude just described?

It does not seem to be self-contradictory to suppose that this might happen. So I suppose we must say that it is not absolutely inconceivable that it should happen. But I am strongly inclined to think that it is, for us as we are, practically inconceivable. The human commitment to participation in ordinary interpersonal relationships is, I think, too thorough going and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as interpersonal relationships as we

normally understand them; and being involved in interpersonal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question (p. 11).

But there is a further implicit point that the paper goes on to draw out. When we normally adopt the objective attitude it is because we view the person who is to be controlled, rather than regarded as a moral agent, as incapacitated. It is not because we view them as 'determined' in the sense of the philosophical thesis of determinism. And in that sense, even if determinism were true it would not justify the objective attitude.

### A niggling doubt?

Now there may remain a niggling doubt that Strawson has answered the wrong question: how we *would* react to the truth of determinism, not how we *should* react to its truth. But he claims to react in this way is to miss the point of his claims so far.

It might be said that all this leaves the real question unanswered, and that we cannot hope to answer it without knowing exactly what the thesis of determinism is. For the real question is not a question about what we actually do, or why we do it. It is not even a question about what we would *in fact* do if a certain theoretical conviction gained acceptance. It is a question about what it would be *rational* to do if determinism were true, a question about the rational justification of ordinary interpersonal attitudes in general. To this I shall reply, first, that such a question could seem real only to one who had utterly failed to grasp the purport of the preceding answer, the fact of our natural human commitment to ordinary interpersonal attitudes. This commitment is part of the general framework of human life, not something that comes up for review as particular cases can come up for review within this general framework. And I shall reply, second, that if we could imagine what we cannot have, namely, a choice in this matter, then we could rationally only in the light of an assessment of the gains and losses to human life, its enrichment or impoverishment; and the truth or falsity of a general thesis of determinism would not bear on the rationality of *this* choice (p. 13).

### What optimists generally miss

The second half of Strawson's (1974) paper is argumentatively less dense. Broadly it goes on to argue that attention to the general framework of direct engaged reactions provides something that is often missing from an optimist's account of moral concepts. They instead base an account on a drier, less engaged account of human practice. Thus by adding in the basis of engaged reactions, Strawson hopes that he can give the pessimists that extra something, which they think must underpin moral judgements and which they construe as a genuine and deep sense of freedom.

This methodological claim, however, seems more a comment on the contingent way the philosophical debate has been conducted than on the underlying issues. Pessimists might

agree that moral judgements run deep in our nature and are based on natural reactions and still call them into question if determinism is true. For that reason it is the argument halfway through the paper, and summarized above, which is the most crucial.

### In summary

Strawson's paper is broadly in the same spirit as the work of Ryle (1963) and Meldon (1961) discussed at the start of this chapter and J.L. Austin from Part 1. It attempts, by more careful description of our (linguistic and other) practices to eschew the need for 'panicky metaphysics' (in this case, a sense of freedom which transcends the causal order). But does he really side-step the problems that Spence's (1996a) paper raises for freedom?

In the second half of his paper, Spence emphasizes the neural underpinnings of conscious free will, arguing that causal determinism is incompatible with there being any genuine sense of freedom of will. We are now in a position to give a response of sorts. We can argue that Spence's claim turns on a false opposition between the kind of freedom that we value and is a presupposition of moral responsibility, and the determinism that is a feature of the causal order. Despite first appearances, freedom in the former sense does not require the falsity of determinism. There is no connection between the discovery of the causal precursors to action and the justification of taking an objective attitude to subjects or agents.

This Strawsonian thought may or may not be combined with a causal theory of action, which would interact with the claims in the first half of Spence's paper. This was summarized earlier as the claim that, although we might think of exercises of free will in action and thought as always conscious, there are both normal and pathological instances where we have thoughts or we act apparently deliberately, but do not experience those acts or thoughts as consciously willed by us. A causal theory of action is an attempt to say what it is for action to be willed: it is for it to be caused (in the right way) by beliefs and desires. According to Davidson, these are in turn states of the body or brain. So one approach to Spence's clinical work is to fit it into just such a general philosophical account.

We have, however, seen that there are problems with arguments for a causal theory of mind in general and Davidson's version of monism in particular. So if instead the right moral is the adoption of a non-causal theory, Spence's results cannot be so accommodated. They would have instead to be taken to be attempting to join together two kinds of description—the mental and the causal—which are incommensurable. A non-causalist has yet to say just what it is for an action to be willed except in a negative way that it was not compelled or some such. Here further clinical work may well be very important. Nevertheless, to repeat the point just made, even the non-causalist account can help itself to Strawson's argument that there is not yet an intolerable threat to freedom raised by the results of brain imaging.

## Reflection on the session and self-test questions

Write down your own reflections on the materials in this session drawing out any points that are particularly significant for you. Then write brief notes about the following:

1. What is the role of reactive attitudes in Strawson's account of the relation of freedom and determinacy?
2. How does the adoption, or not, of an objective attitude towards agents dovetail with the suspension of reactive attitudes like resentment?
3. What argument does he offer against the universal adoption of an objective outlook?

## Reading guide

Useful collections concerning causal theories of action include:

- ◆ Alfred Mele (ed.) (1997) *The Philosophy of Action*.
- ◆ John Heil and Alfred Mele (ed.) (1993) *Mental Causation*.
- ◆ White's (1968) collection, *The Philosophy of Action*, includes Austin's (1956/7) *A Plea for Excuses* (see Part I) and Anscombe's (1959) *Intention*.
- ◆ Austin's (1966) *Three Ways of Spilling Ink*, Hampshire's (1959) *Thought and Action*, and Winch's (1972) *Ethics and Action*, are important classics.
- ◆ Jennifer Hornsby's (1980) *Actions* is more difficult but captures recent trends.

(See also Reading guide to Chapter 6.)

## Non-causal theories of action

The view of willing that inspired many non-causal theories of action can be found in: Wittgenstein (1953) *Philosophical Investigations*, §611–630.

It is supported in:

- ◆ G.E.M. Anscombe (1959) 'Intention' (in *Proceedings of the Aristotelian Society*, and later reprinted in White's (ed.) (1968) *The Philosophy of Action*)
- ◆ Malcolm, N. (1968) The conceivability of mechanism *Philosophical Review*
- ◆ Melden, A.I. (1961) *Free Action*
- ◆ P. Winch's (1960) *The Idea of Social Science and its Relation to Philosophy* (2nd edn)

## Causal theories of action

- ◆ Davidson's (1980) causal theory of action is set out in a number of essays collected in his *Essays on Actions and Events*.
- ◆ A useful introduction is Evnine's (1991) *Donald Davidson*.

- ◆ More specific work on Davidson's account can be found in Child's (1994) *Causality Interpretation and the Mind*, especially chapter 3, and in many of the essays in LePore and McLaughlin (1985) *Essays on Actions and Events*.

## Freedom and determinism

The philosophical debate about freedom and determinism is set out in the following:

- ◆ Daniel C. Dennett (1984) *Elbow Room* and (2003) *Freedom Evolves*
- ◆ Ilham Dilman (1999) *Free Will: an historical and philosophical introduction*
- ◆ Robert Kane (ed.) (2001) *Free Will* (1996) The significance of free will
- ◆ Gary Watson (ed.) (2003) *Free Will*.

## Applied work

- ◆ Spence's article: Spence (1996a) 'Free will in the light of neuropsychiatry' (*Philosophy, Psychiatry, & Psychology*, pp. 75–90); has commentaries by Frith (1996, pp. 91–94), Libet (1996, pp. 95–96), Stephens (1996, pp. 97–98), and reply by Spence (1996b, pp. 99–100).
- ◆ Spence's work is further developed in Spence (2001) 'Alien control: from phenomenology to cognitive neurobiology' (*Philosophy, Psychiatry, & Psychology*, pp. 163–172); with commentary by Pacherie (2001, pp. 173–176).
- ◆ Further applied work on free will includes: Waller (2004a) 'Neglected psychological elements of free will' (*Philosophy, Psychiatry, & Psychology*, pp. 111–118); with commentary by Lieberman (2004, pp. 119–124), and reply Waller (2004b, pp. 125–128).
- ◆ The Harvard psychologist Daniel Wegner sets out, in *The Illusion of Conscious Will* (2002), to deconstruct the idea that we consciously will our actions using experimental psychology.
- ◆ Stephens G.L. and Graham, G. (1996). Psychopathology, freedom and the experience of externality *Philosophical Topics* is an examination of just what our capacity to be alienated from our own states tells about us the importance of ownership of states for freedom & action.

## References

- Anscombe, G.E.M. (1959). Intention. *Proceedings of the Aristotelian Society* 57: 321–332. (Reprinted in White, A.R. (ed.) (1968). *The Philosophy of Action*. Oxford: Oxford University Press.)

- Austin, J.L. (1956/7). A plea for excuses. *Proceedings of the Aristotelian Society*, 57: 1–30. (Reprinted in White, A.R. (ed.) (1968). *The Philosophy of Action*. Oxford: Oxford University Press, pp. 19–42.)
- Austin, J.L. (1966). Three ways of spilling ink. *Philosophical Review*, 75: 427–440. (Reprinted in White, A.R. (ed.) (1968). *The Philosophy of Action*. Oxford: Oxford University Press.)
- Bishop, J. (1983). Agent-causation. *Mind*, 92: 61–79.
- Child, W. (1994). *Causality Interpretation and the Mind*. Oxford: Oxford University Press.
- Davidson, D. (1980). *Essays on Actions and Events*. Oxford: Oxford University Press.
- Davidson, D. (1982). Paradoxes of irrationality. In *Philosophical Essays on Freud* (ed. J. Hopkins and R. Wollheim). Cambridge: Cambridge University Press, pp. 289–305.
- Dennett, D.C. (1984). *Elbow Room*. Cambridge, MA: MIT Press.
- Dennett, D.C. (2003). *Freedom Evolves*. New York: Viking.
- Dilman, I. (1999). *Free Will: an historical and philosophical introduction*. London: Routledge.
- Evnine, S. (1991). *Donald Davidson*. Oxford: Polity.
- Frith, C. (1996). Free will in the light of neuropsychiatry. (Commentary on Spence, 1996a) *Philosophy, Psychiatry, & Psychology*, 3(2): 91–94.
- Hampshire, S. (1959). *Thought and Action*. London: Chatto and Windus. (Reprinted in White, A.R. (ed.) (1968). *The Philosophy of Action*. Oxford: Oxford University Press.)
- Heil, J. and Mele, A. (ed.) (1993). *Mental Causation*. Oxford: Oxford University Press.
- Hornsby, J. (1980). *Actions*. London: Routledge & Kegan Paul.
- Kane, R. (1996). *The Significance of Free Will*. Oxford: OUP.
- Kane, R. (ed.) (2001). *Free Will*. Oxford: Blackwell.
- LePore, E. and McLaughlin, B. (1985). *Actions and Events*. Oxford: Blackwell.
- Libet, B. (1996). Free will in the light of neuropsychiatry. (Commentary on Spence, 1996a) *Philosophy, Psychiatry, & Psychology*, 3(2): 95–96.
- Lieberman, P.B. (2004). Action, belief, and empowerment. (Commentary on Waller, 2004a) *Philosophy, Psychiatry, & Psychology*, 11(2): 119–124.
- Malcolm, N. (1968). The conceivability of mechanism. *Philosophical Review*, 77: 45–72.
- McDowell, J. (1994). *Mind and World*. Cambridge, MA.: Harvard University Press.
- Melden, A.I. (1961). *Free Action*. London: Routledge.
- Mele, A.R. (1997). *The Philosophy of Action*. Oxford: Oxford University Press.
- Pacherie, E. (2001). Agency lost and found: a commentary on Spence. (Commentary on Spence, 2001) *Philosophy, Psychiatry, & Psychology*, 8/2/3: 173–176.
- Ryle, G. (1963). *The Concept of Mind*. London: Penguin.
- Spence, S.A. (1996a). Free will in the light of neuropsychiatry. (Commentaries by Frith, C. (1996, pp. 91–94), Libet, B. (1996, pp. 95–96), and Stephens, G.L. (1996, pp. 97–98) and reply by Spence, S.A. (1996b, pp. 99–100)) *Philosophy, Psychiatry, & Psychology*, 3(2): 75–90.
- Spence, S.A. (1996b). Response to the Commentaries (on Spence, 1996a). *Philosophy, Psychiatry, & Psychology*, 3(2): 99–100.
- Spence, S. (2001). Alien control: from phenomenology to cognitive neurobiology. (with commentary by Pacherie, E. (2001, pp. 173–176)) *Philosophy, Psychiatry, & Psychology*, 8/2/3: 163–172.
- Stephens, G.L. and Graham, G. (1996). Psychopathology, freedom and the experience of externality *Philosophical Topics*, 24: 159–182.
- Strawson, P.F. (1974). Freedom and resentment. In *Freedom and Resentment, and Other Essays*. London: Methuen, pp. 1–25.
- Tanney, J. (1995). Why reasons may not be causes. *Mind and Language*, 10: 105–128.
- Waller, B.N. (2004a). Neglected psychological elements of free will. (Commentary by Lieberman, P.B. (2004, pp. 119–124) and reply Waller, B.N. (2004b, 125–128)) *Philosophy, Psychiatry, & Psychology*, 11(2): 111–118.
- Waller, B.N. (2004b). Comparing psychoanalytic and cognitive-behavioral perspectives on control. (Response to Lieberman, 2004) *Philosophy, Psychiatry, & Psychology*, 11(2): 125–128.
- Watson, G. (ed.) (2003). *Free Will*. Oxford: Oxford University Press.
- White, A.R. (ed.) (1968). *The Philosophy of Action*. Oxford: Oxford University Press.
- Winch, P. (1960). *The Idea of Social Science and its Relation to Philosophy*, (2nd edn). London: Routledge & Kegan Paul.
- Winch, P. (1972). *Ethics and Action*. London: Routledge and Kegan Paul.
- Wittgenstein, L. (1953). *Philosophical Investigations*. Oxford: Basil Blackwell.