

INTENTION, AGENCY AND CRIMINAL LIABILITY:

*Philosophy of Action and
the Criminal Law*

R. A. Duff

Basil Blackwell

Contents

Preface	viii
Table of Cases	x
Table of Statutes	xiii
Abbreviations	xiv
1 INTRODUCTION	1
1.1 Cases and Questions	1
1.2 <i>Actus Reus</i> and <i>Mens Rea</i>	7
PART I INTENTION AND AGENCY	
2 LEGAL CONCEPTIONS OF INTENTION	15
2.1 The Meaning of Intention	15
2.2 Proving Intention	27
2.3 Why Define Intention?	31
3 INTENTION IN ACTION – A PARADIGM	38
3.1 Preliminaries	38
3.2 Intention, Bare Intention and Decision	44
3.3 Intention and Reasons for Action	47
3.4 Intention and Desire I	52
3.5 Intention, Desire and Belief I	55
3.6 Intention, Desire and Belief II	58
3.7 Intention, Success and Causation	63
3.8 Intention and Desire II	66
4 INTENTION, FORESIGHT AND RESPONSIBILITY	74
4.1 Direct and Oblique Intention	74
4.2 Intentional Action and Responsibility	76
4.3 Aspects of Responsibility	82

4.4	Intention and Circumstances	87
4.5	Individuating Effects	89
4.6	Intentional Agency and Probable Consequences	95
5	COMPETING CONCEPTIONS OF AGENCY	99
5.1	Intention and Responsibility	99
5.2	A Consequentialist View of Responsible Agency	105
5.3	A Non-consequentialist View	111
6	INTENTION, ACTION AND STATES OF MIND	116
6.1	Dualism and the Mental Element in Crime	116
6.2	The Argument from Analogy	120
6.3	Actions and 'Colourless Movements'	123
6.4	Identifying Mental States	127
6.5	An Alternative View	129
PART II SUBJECTIVE AND OBJECTIVE		
7	RECKLESSNESS	139
7.1	Extending the Paradigms	139
7.2	Recklessness in the Criminal Law	142
7.3	'Subjectivism' and 'Objectivism'	149
7.4	'The Thought Never Crossed My Mind'	157
7.5	'I Thought She Was Consenting'	167
7.6	Implied Malice and Murder	173
8	CRIMINAL ATTEMPTS	180
8.1	Introduction	180
8.2	The Significance of Failure	184
8.3	The <i>Mens Rea</i> of Attempts I: Subjectivism and the Current Law	192
8.4	The <i>Mens Rea</i> of Attempts II: Why Attempts should be Intended	199
8.5	Concluding Remarks	206
	Bibliography	207
	Index	215

To HGM and VJM

6

*Intention, Action and States of Mind*6.1 *Dualism and the Mental Element in Crime*

Intention is a 'state of mind with which a person acts' (1989 Code, cl. 6). But what is a 'state of mind', and how are such states related to action? These philosophical issues are not prominent in the law, but they underlie the legal question of how intention can be proved: for any answer to that question presupposes some view of the nature of mental states, and of their relation to the conduct through which they are typically revealed.

Indeed, we found such a view in juristic comments on the proof of intention and on the presumption that agents intend the 'natural and probable consequences' of their acts (see pp. 28–31 above): for many jurists assume a *dualist* view of the mind, portraying intentions as private mental states or occurrences which must be inferred from external behaviour. Now an adequate discussion of Dualism would require a book to itself, but we should briefly consider its implications, and its adequacy, for our understanding of intention and action.¹

Classical Dualism holds that human beings consist of two distinct elements: a physical body, which occupies and moves in space, and a non-physical mind, which thinks and feels. Bodies are public, whereas minds are essentially private: others can directly observe my body and its movements, but only I can directly observe what is going on in my mind.²

1 On Dualism, see P. Smith and O.R. Jones, *The Philosophy of Mind*, Part I; P. Carruthers, *Introducing Persons*; I. Dilman, *Matter and Mind*, Part II; G. Ryle, *The Concept of Mind*. This chapter owes much to R. Shiner, 'Intoxication and responsibility'.

2 See R. Descartes, *Meditations on First Philosophy*; J. Cottingham, *Descartes*, ch. 5.

This does not mean that we can never know what another person is thinking or feeling: for mind and body are connected and act on each other. The pain I feel when I burn my arm involves a mental sensation which is caused by a physical process, and which itself causes such physical behaviour as wincing or crying out; and when I move my hand, that physical movement has a mental cause. Such connections between mind and body enable us to infer the private mental states of others from their physical behaviour: but they are connections between two distinct entities; and we can come to know of another's mental states *only* by such inferences from her observable behaviour to the mental states which we suppose to be connected to it.

A defendant's intentions must, therefore, always be inferred from 'external' or 'circumstantial' evidence. For 'intention is a state of mind', and we 'can never look into the mind of an accused person' or 'actually see what his intent was': we must, assisted by presumptions such as that agents intend the 'natural and probable consequences' of their acts, infer his intentions from such evidence as we have – from 'what he did, what he said and all the circumstances of the case' (see pp. 28–9 above).

Human actions, on this view, consist of two distinct elements: an external, observable bodily movement; and an internal, unobservable mental state of intention. This has an obvious affinity to the legal distinction between *actus reus* and *mens rea*: the *actus reus* consists, we might say, in the 'external elements' of the offence which others can observe; the *mens rea* in the 'internal' mental states which must rather be inferred. But the legal distinction between *actus reus* and *mens rea*, while it may reflect dualist assumptions, does not exactly match the dualist's distinction between mind and body.

First, the *actus reus* itself must sometimes be defined in terms of the defendant's own mental state. To take just one example, the *actus reus* of a criminal attempt involves an act which is 'more than merely preparatory to the commission of the offence': but what makes an act 'more than merely preparatory' is the agent's intention to commit the relevant offence; the act of putting sugar in my aunt's tea constitutes the *actus reus* of attempted murder only given that I believe it to be poison and intend to kill her.³

Second, the *actus reus* is anyway often said to include (normally) a mental element: for it usually includes a 'voluntary act', and what makes an act voluntary is a mental element of 'will' or 'volition'. A bodily movement as such (a movement of my arm) does not constitute a voluntary act, since

3 Criminal Attempts Act 1981, s. 1(1); see, more generally, A.C.E. Lynch, 'The mental element in the *actus reus*'.

it could be either voluntary or involuntary (I might move my arm; or it might jerk involuntarily): a voluntary act is, on one common account, a *willed* bodily movement – a movement caused by a mental act of volition. I shall not discuss this account of voluntary agency here: we need note only that, although it draws the dualist's distinction between mind and body (between the bodily movement itself and the mental act of 'willing'), it does not distinguish *actus reus* from *mens rea* precisely in line with that distinction; for the *actus reus* now includes an element of mind as well as of body.⁴

We should note one further preliminary point before attending to the attractions and defects of Dualism.

Classical Dualism was a *metaphysical* doctrine about what kinds of thing exist; minds exist, it held, as non-physical entities. Now materialists reject that metaphysical doctrine, arguing that human beings are purely physical beings and that minds are identical with brains: but they may still be *epistemological* dualists, who share the classical dualist's view of how we can come to *know* another's mental states. For they may say that, though we could *in principle* directly observe another's mind by observing her brain, we cannot *in fact* do so. They too may thus distinguish the external behaviour which we can directly observe from the inner mental states (or brain states) which we cannot directly observe, and say that we can in fact come to know the mental states of others only by inferring them from the behaviour which we can directly observe; and they are then epistemological, but not metaphysical, dualists. Judges who note the lack of a 'meter' or 'X-ray machine' which would show us the agent's intentions (see pp. 28–9 above) might be materialists (or metaphysical dualists with a vivid turn of phrase): but they are epistemological dualists who hold that we must always infer the mental states of others from their external behaviour. The objections to Dualism that I shall discuss are aimed at metaphysical Dualism: but they can be adapted to apply to epistemological versions of Dualism as well.

Dualism may seem to be an obvious implication of our ordinary way of thinking about the mind: surely we do draw the dualist distinction between public bodies and private minds – between the external bodily movements of others, which we can directly observe, and their inner mental states, which we cannot directly observe.

4 See *S&H*, pp. 39–41; *TCL*, pp. 147–8; *C&K*, pp. 86–9; H.L.A. Hart, 'Acts of will and responsibility'.

Surely, for instance, I can see another's bodily movements, while being unsure whether they are intentional or what he intends: I see Ian's hand strike Pat's chin, but am unsure whether he meant to move his hand (perhaps it was an involuntary spasm) or whether he meant to hit Pat (perhaps he did not see her); or perhaps I mistakenly suppose that he intended to hit Pat, when in fact the blow was involuntary or accidental. Familiar cases like this surely require us to distinguish the bodies and bodily movements which we can observe, from the minds and mental states which we cannot directly observe, but must infer. I am neither unsure nor mistaken about Ian's external behaviour: but I am unsure or mistaken about the mental states which I try to infer from that behaviour.

An agent can also deceive us about her intentions or hide them from us. I can, of course, sometimes conceal my bodily movements: I can close the door, so that no one can see what I am doing; you must then infer my movements from such circumstantial evidence as is available. But the impossibility of direct observation in this case is merely *contingent*: you could in principle see what I am doing by opening the door; you know what it would be to observe my movements directly. My intentions, however, are *necessarily* hidden unless and until I choose to reveal them by announcing them, or by acting in a way which makes them clear; they cannot even in principle be directly observed by others. This is most obviously true of bare intentions, which I have not yet put into action: these are, surely, pure mental states without external behaviour; you can know what they are only if I choose to reveal them. This again supports the dualist distinction between public bodies and private minds: my body and its movements may be contingently hidden from others, but my intentions and other mental states are necessarily hidden from their direct observation.

Others must infer my mental states from my observable behaviour, and may draw mistaken inferences: but I surely know directly, with no need for inferences and no room for mistakes, what I think or intend; and this privileged, direct awareness which we each have of our own mental states again supports the dualist distinction. While my body and its movements are open to direct observation by others, my mind is a private, inner realm to which I alone have direct access.

Despite its apparent plausibility, however, I think that Dualism is a deeply mistaken doctrine: it distorts what it seeks to explain and has vitiated attempts to provide an adequate account of intention in the law. I shall indicate some of the main objections to it, and outline an alternative account of the nature of intention as a 'state of mind' and of the relation between intention and action.

6.2 The Argument from Analogy

I shall focus on three central features of a simple dualist view. The first concerns the reliability of those inferences from external behaviour to inner mental states on which, the dualist claims, our knowledge of another's mind always depends.⁵

These inferences are from *observed* behaviour to an *unobserved* mental state. Now inferences from the observed to the unobserved are usually based on and justified by an *observed* and regular correlation between them. My doctor's inference from the rash on my body to the diagnosis that I have measles is justified by the previously discovered regular correlation between rashes like this and the relevant virus: she can infer the presence of that presently unobserved virus only because it *has* been observed to be regularly correlated with the kind of rash that she now observes. But our inferences from another's behaviour to his mental states cannot be based on correlations which we have observed between the behaviour of others and *their* mental states: for we can never directly observe the mental states of others. The only case in which we can observe correlations between external behaviour and inner mental states is our own. I am directly aware of my own mental states, and can observe correlations between them and my external behaviour and situation; these observed correlations must provide the basis for my inferences from the behaviour of others to their mental states.

This is the Argument from Analogy. I see bodies around me, which resemble mine and behave in ways similar to mine. I know that my body is connected to a mind. So I infer, by analogy with my own case, that these other bodies are also connected to minds. I observe in my own case a correlation between an external stimulus (contact with fire) and a mental state (pain), or between that mental state and a certain kind of behaviour (wincing or crying out): so when I observe another body which resembles mine being subjected to a similar stimulus, or exhibiting similar behaviour, I infer, by analogy with my own case, that a similar mental state has occurred in the mind which I suppose to be connected to that body.

But can such inferences be reliable? Inferences from the observed to the unobserved are usually thought reliable only if they are based on a large number of cases in which the relevant correlation has been observed: we

5 See P. Carruthers, *Introducing Persons*, pp. 7–21; N. Malcolm, 'Knowledge of other minds'.

should reject inferences based on what has been observed in just one case ('I know that this man will attack me because he is red-haired and the only other red-haired man I met attacked me'). How then can I properly base my claim to know what another is thinking or feeling on the correlations between behaviour and mental state which I have observed in just one case – my own? I have, of course, observed a large number of such correlations in my own case. But they are still correlations involving only one body; and I surely cannot reliably assume that what is true of this one body is true of all these others (the fact that one red-haired man has frequently attacked me does not make the inference that any red-haired man will attack me sound).

Furthermore, inferences from the observed to the unobserved are, normally, inferences which could *in principle* be tested by observing the currently unobserved item whose presence we infer. The doctor who infers that I have measles could, given the right equipment, directly observe the virus whose presence she infers; and what underpins her claim that that virus is present is the well-founded belief that she *would* observe it directly if she undertook the appropriate procedure. For the dualist, however, minds are *necessarily* private to their owners: I cannot even in principle directly test the inferences which I make to another's mental states, since there is no procedure through which I could observe directly the mental state whose presence I now infer in his mind. But how can I place any trust in inferences which cannot even in principle be tested?

A materialist might claim that we *could*, in principle, test our inferences to the mental states of others: for we could in principle directly observe their mental states, by observing their brains. But how do we know which brain states constitute which mental states? We must find that certain brain states occur when certain mental states occur, thus identifying those brain states with those mental states: but to do this we must be able to know that a relevant mental state is occurring, independently of observing the brain state with which we then identify it; and this confronts us with the same problem as the metaphysical dualist. To find that a certain brain state constitutes a feeling of pain, for example, we must find that that brain state occurs whenever someone feels pain. To do this we must be able to know that someone is in pain when her brain is in that state: but if we ask how we can come to know this, the epistemological dualist must give the same answer as a metaphysical dualist. If we could establish the appropriate identifications of brain states with mental states, we could then discover that someone is in pain by observing her brain states: but in order to establish those identifications we must be able to discover that others are in pain by a method which does not rely on observing their brain states; and for both metaphysical and epistemological dualists, that method must

involve the Argument from Analogy. The objection that that argument is unsound thus undermines both metaphysical and epistemological Dualism.

Could we defend the Argument from Analogy by saying that there is one way in which we can directly know, and thus directly verify our claims about, another's mental states, since she can *tell* us what her mental states are? Jurists seem sometimes to take this view.

Intention may be directly proved from what the defendant says. Evidence may be given of what he said contemporaneously with the act ... or of his prior or subsequent admission of what he intended to do. ... If the defendant does not give the court this assistance, the jury ... will have no direct access to his mind. (*TCL*, p. 80)

Similarly, in a different jurisdiction,

There was no evidence of any express statement made by the prisoner as to his intention when he inflicted the fatal blows upon his wife, and ... he gave no evidence at the trial. The necessary intention must therefore be a matter to be inferred from his actions and statements. (*Foy*, p. 233)

In the absence of an 'express statement' by the agent, we must infer his intentions from such evidence as we have: but, it seems, those intentions are directly *revealed*, without need of inference, in any 'express statement ... as to his intention'. So could we not say, in general, that a person's own descriptions of her mental states give us direct, non-inferential knowledge of those states, and thus give us a way of checking our claims about them?

But this will not do. According to Dualism, we directly observe only bodies and the movements and sounds which they make. If someone speaks to me I directly observe only certain movements and sounds. To know that a *person* is *telling* me something (that these are not just meaningless sounds emanating from a mindless body), I must know that this is a person who intends to communicate with me – that these sounds are caused by a particular mental state; and I can know this only by making an analogical inference from the sounds which I hear emerging from this body to the existence and the intentions of a mind which caused them – to their status and meaning as speech. The claim that I must infer others' mental states from the physical behaviour which I directly observe applies to speech as to any other kind of behaviour: a dualist cannot consistently make an exception of speech and claim that this kind of behaviour, uniquely, gives us direct access to another's mind (see pp. 123–6 below).

Dualists are not, of course, silenced at once by such objections to the Argument from Analogy: there are further moves which they can make to defend that argument – and further moves which the critic can then make to try to show that the defence fails. But we cannot pursue this discussion

here. We should note only that if the inferences which Dualism requires us to make from external behaviour to inner mental states cannot be sustained, then Dualism must lead us to scepticism about the very existence, let alone the contents, of other minds. For if the Argument from Analogy fails; if I cannot properly infer the existence or the contents of other minds from my knowledge of my own mental states and their relation to my physical behaviour, together with the resemblances which I observe between my body's behaviour and that of other bodies: then I cannot claim to know even that other people exist (that these other bodies are related to minds), let alone what their mental states are.

6.3 Actions and 'Colourless Movements'

The second feature of Dualism to which we should attend concerns the 'external behaviour' from which we must infer the mental states of others.

What we directly observe, Dualism holds, are not *people* and their *actions*, but bodies and their 'colourless movements'. For people have minds, and actions involve intentions: but these are hidden from our observation and must be inferred; in advance of such inferences, we cannot know that the bodies we see are people's bodies, or that their movements constitute actions:

If I look out of the window and see men crossing the square ... I normally say that I see the men themselves Yet do I see any more than hats and coats which could conceal automatons? I *judge* that they are men. (R. Descartes, *Meditations* II, p. 21)

Dualist judges who say that the defendant's intentions must always be inferred from what he 'did' and 'said' should therefore more strictly say that they must be inferred from the movements of his body and the sounds it emitted: for to count such movements as actions (as 'what he did'), or such sounds as speech (as 'what he said'), is to go beyond what Dualism holds that we can directly observe; it assumes a 'mental element' which we cannot observe but must infer.

Our descriptions of the observed behaviour of others are not, of course, usually couched in the bare language of bodies and colourless movements. I say, 'I saw Ian hit Pat on the chin', not 'I saw a hand move and come into contact with a face'; or 'I heard Ian say what he had done', not 'I heard a body emit certain sounds': I ascribe to a *person* (not just a body) an intended *action* of hitting or talking (not just a bodily movement or emission of sound). The discussion of what Mr Moloney could have intended likewise began from the fact, not just that his finger moved, but that he pulled the trigger of a gun; and the discussion in *Hancock and*

Shankland began from the fact, not merely that their bodies moved in certain ways, but that they pushed a block of concrete off a bridge. But such descriptions cannot, on a dualist view, be descriptions of what was directly observable: they rather embody complex sets of inferences from the bodily movements which could be observed to the mental states in virtue of which those movements constituted human actions.

This account of what is given in observation is common to Dualism and its familiar opponent, Behaviourism. Both insist that we directly observe only physical bodies and their colourless movements, and that 'mental states' must be inferred from those physical phenomena. They disagree about the nature of such 'mental states', and about how they are related to physical behaviour. For a dualist, physical behaviour is the evidence from which we make inferences to the hidden realm of the mind: for a behaviourist, there is no such hidden realm – talk of mental states is nothing more than talk of patterns of behaviour.⁶

To say that A intends to do X is, for a dualist, to infer some distinct and hidden mental state from A's observable behaviour. For a behaviourist, it is rather to predict her observable behaviour; and what founds that prediction is not an inference to a hidden mind, but the observable patterns of A's behaviour and the correlations between it and certain external stimuli. I see Ian moving down the road, and say that he intends to go shopping. For a dualist, my claim concerns the hidden contents of his mind: it generates a prediction about his future behaviour (entering shops and buying goods), and may be based on what I know of his past behaviour (he always shops on Tuesdays) or his reasons for action (he has run out of food); but it asserts the existence of a distinct mental state, which is caused by that reason for action and will cause that future behaviour. For a behaviourist, my claim concerns *only* the observable pattern of his present and future behaviour: its meaning consists solely in the prediction that he will make certain movements, and it is founded solely on the correlations which I suppose to hold between such behaviour and its being Tuesday, or his having run out of food; it posits no hidden mental states which connect such stimuli to his observable behaviour.

Part of what motivates Behaviourism is the dubious status of the dualist's inferences to hidden mental processes: a truly 'scientific' account of

6 On Behaviourism (and the different forms which it can take), see P. Carruthers, *Introducing Persons*, ch. 4; P. Smith and O.R. Jones, *The Philosophy of Mind*, pp. 144–6; N. Malcolm, *Problems of Mind*, pp. 91–102; I. Dilman, *Matter and Mind*, Part II, chs 8–9; M.S. Moore, *Law and Psychiatry*, pp. 36–9. The phrase 'colourless movements' comes from a behaviourist: C. Hull, *Principles of Behaviour*, p. 25.

human behaviour has no room for such mysteries. Part of what motivates Dualism, on the other hand, is the belief that there is more to human action than mere patterns of bodily movement: surely we cannot capture the richness of human actions merely by describing observable bodily movements? Now if we accept the assumption which Dualism and Behaviourism share, that we directly observe only bodies and their movements, we must accept one of these views: for actions must then consist either (as behaviourists claim) purely in patterns of bodily movement, or (as dualists claim) in bodily movements plus some further, mental ingredient; and people must either be nothing more or other than physical bodies, or bodies plus distinct minds. But should we accept that assumption?

What can a dualist make of the fact that our descriptions of the observable behaviour of others do not normally just describe bodies and their colourless movements?

She may reply that our ordinary descriptions go beyond the bare facts which we directly observe; they embody inferences which we make (perhaps unconsciously) from what we observe to the agent's mental states. If I describe a mark on Ian's skin as a burn, my description embodies the inference that that mark was caused by heat: if I am experienced in such matters, I may not go through a conscious process of inferential reasoning ('this mark resembles others which I found were caused by heat ...'); but my description none the less embodies an inference from the observed mark to its unobserved cause. Similarly, our ordinary descriptions of behaviour include more than the mere movements of bodies because they embody inferences which we make from the bodily movements which we observe to the unobserved mental states which we think caused them: inferences which we are so used to making that we do not make them explicit to ourselves. When I say, 'I saw Pat driving her car', I am not strictly describing what I *observed*: my description embodies the inferences which I (perhaps unconsciously) made from the bodily movements which I actually saw to the mental states of intention and knowledge which caused them.

Now if this is true, we should be able to provide a purified description of what we directly observe, and an explanation of the inferences upon which our ordinary descriptions of behaviour depend. If my description of Pat's behaviour as 'attacking Ian' embodies the inferences I have made from the bodily movements which I directly observe to certain hidden mental states, I should surely be able to make those inferences explicit, and to describe the bare bodily movements on which they are based. But is this possible?

I could no doubt try to see, and describe, Pat's behaviour in the

appropriate terms, as a set of bodily movements which neither involve nor imply any 'mental element' of intention. It would, in fact, be very difficult to do this: but even if I can do it, I am surely not making explicit the descriptions of colourless movements from which, on the dualist account, my descriptions of her actions were inferred. I am rather taking a new and different perspective on her behaviour; I try to see it not, as I naturally see it, as action, but as mere bodily movement. In other words, to describe her behaviour merely as bodily movements involves not, as the dualist must claim, stripping away the inferential accretions which my ordinary descriptions embody, to lay bare the evidence from which I infer those descriptions; but rather abandoning such ordinary descriptions, and the perspective from which they are offered, altogether (see F. Ebersole, 'Where the action is').

Furthermore (to relate this argument to that of the last section), even if I could provide an austere description of Pat's behaviour as a set of mere bodily movements, it is hard to see how this could form a basis for reliable inferences to her hidden mental states. We quite readily infer people's intentions from their *actions*; I can easily infer that Ian intends to go to London from the fact that he buys a rail ticket to London. The basis of that inference, however, is not a set of mere bodily movements, but his action of buying a ticket; and from a description of his behaviour merely as a set of bodily movements I do not think that I could infer anything at all about his intentions. Once we see just how bare our descriptions of behaviour must be, if they are to describe only the colourless bodily movements which are all that we can, according to the dualist, directly observe, we must also see that such descriptions are utterly inadequate as a basis for the kinds of inference which, according to the dualist, we must make.

This should undermine the dualist claim that we directly observe only bodies and their movements; that we must infer from those bodies and movements the 'mental elements' which turn them into people and actions. The oddity of that claim is concealed if we say that what we directly observe is 'behaviour': for we naturally take 'behaviour' to involve actions rather than mere bodily movements (I describe Ian's 'behaviour' as 'hitting Pat', or 'buying a ticket'). But on a strict dualist view the 'behaviour' which we can directly observe consists only in colourless bodily movements; and I have suggested that this is neither a possible account of what we observe when we observe other people, nor a possible basis for inferences to their intentions or other mental states. If this suggestion is right, it rebuts both Dualism and strict Behaviourism, by rebutting the assumption which they share: but before supporting it by an account of what we do directly observe, we must look briefly at another aspect of Dualism.

6.4 Identifying Mental States

Our knowledge of other minds involves, Dualism claims, inferences from colourless bodily movements to distinct, hidden mental states or processes. But we do not, I have argued, start from colourless bodily movements; and we must ask now whether we make inferences to distinct and hidden mental states.⁷

The dualist distinguishes mind from body, inner mental state from external behaviour, and holds that the former must be inferred from the latter. It follows from this that we must be able to identify our mental states independently of the external behaviour from which they are inferred by others: if A is distinct, and must be inferred, from B, we must be able to identify A independently of B. If intention is a mental state which is distinct, and must be inferred, from external behaviour, I must be able to identify certain mental states of mine as intentions, without reference to the kinds of behaviour which they cause; and in ascribing intentions to others, I am saying that such independently identifiable states exist in their minds. But can we identify intentions independently of behaviour?

I have already noted two objections to the claim that intentions are mental states distinct from external behaviour (pp. 44–7 above). First, intended actions are not always preceded (or accompanied) by any conscious act or state of intending. If a dualist replies that there must in such cases be an *unconscious* prior process of decision or intention-formation, we should ask what justifies this 'must': we may suspect that it is question-begging (since Dualism is true, this must be what happens); and we should note that I can know what I intend to do when I act without having to *discover* something which is, on this account, hidden from my immediate consciousness. Second, we cannot identify, among the conscious thoughts which may precede or accompany an action, some occurrent thought or mental act which can be recognized *in itself* as being an intention or decision: even if we can find some thought of an appropriate form (such as 'I will do X'), we must ask what identifies it as a decision or intention – rather than, for instance, a merely idle thought about the action; and there are no *intrinsic* features of thoughts, when examined independently of the actions to which they may be related, which could distinguish genuine intentions from merely idle thoughts.

7 See N. Malcolm, 'The conceivability of mechanism'; A.I. Melden, *Free Action*, chs VIII–XIII; P. Smith and O.R. Jones, *The Philosophy of Mind*, chs IX, XVII; A.R. White, *Grounds of Liability*, pp. 66–72.

These objections suggest that we cannot separate intention from behaviour as the dualist does; that we can identify our own or other people's intentions only in and through the actions which we or they intend to do. The dualist might reply, however, that we do sometimes separate intention from action: we ascribe bare intentions to others (and know our own bare intentions) before those intentions are put into effect; when an action misfires or goes wrong, we naturally say that the agent's actual behaviour did not accord with her intentions. Does this not show that intention and behaviour *are* distinct, and can be identified independently of each other?

Such examples show that we do *sometimes* identify intention in the absence of the intended behaviour; that intention can to that extent be distinguished from behaviour. But the dualist distinction between intention and behaviour is sharper than this: for it is a distinction between two elements which are related only *contingently*. If I infer another's intentions from her observable behaviour I am, according to the dualist, inferring the existence of a distinct mental state which causes that behaviour; and that inference is based on my discovery of the correlations which obtain between my own inner mental states and my external behaviour. Now such inferences from effect to cause, and such discoveries of correlations between mental state and behaviour, presuppose that mental state and behaviour are logically unrelated; that the existence of the one in no way *entails* the existence of the other: but intention and behaviour cannot be thus logically distinct. If they were thus distinct, it would be conceivable that intentions should *never* produce the behaviour which is intended; that no one should *ever* do what they intend: but this is not conceivable; it cannot be coherently imagined.

Although people do not *always* do, or succeed in doing, what they intend to do, so that 'A intends to do X' does not simply entail that 'A does X', it is not simply a contingent fact that people *usually* do what they intend to do; it is part of the *logic* of intention that if I intend to do X, I will do it unless something (a change of mind or an external obstacle) intervenes. It is not conceivable that people should *never* do what they intend; for no state of mind which was thus unrelated to behaviour could count as intention. But this then shows that intentions cannot be identified, as Dualism requires them to be, independently of behaviour: I do not *discover* correlations between my intentions and my behaviour, or make inferences from the behaviour of others to mental states of intention which are *contingently* related to that behaviour; for intentions cannot be identified as intentions (as distinct, for example, from idle thoughts about an action) except by reference to the actual behaviour which is intended.

If this argument succeeds it shows that intention is not a 'state of mind' which is logically quite distinct and separate from external behaviour; and

thus that the distinction which Dualism draws between 'mind' and 'behaviour' cannot be drawn. An account of what a person intends is not an account of what is happening in some hidden mental realm which is logically unrelated to her observable behaviour: for to say that she intends to do something is to talk about what she will *do* (unless something intervenes to prevent her doing it).

6.5 An Alternative View

I do not claim to have refuted Dualism here: I have only sketched three objections which, if they cannot be met, will show Dualism to be untenable. Rather than discuss the responses which dualists might make to these objections, however, I shall outline an alternative view of intention and action which rejects the basic assumption, shared by dualists and behaviourists, that we directly observe only bodies and their colourless movements.

If we accept that assumption, we must choose between Dualism and Behaviourism: for actions must then consist either simply in patterns of bodily movement, or in bodily movements plus some extra (mental) ingredient. But neither of these accounts seems tenable. Dualism requires us to make inferences from the colourless movements which we supposedly observe to a hidden mental realm: but such inferences seem impossible. Behaviourism rejects such inferences to a hidden mental realm: but it seems impossible that any descriptions of mere bodily movements, however complex, can capture the meanings of our ordinary action-descriptions. Dualists rightly insist, and behaviourists wrongly deny, that there is more to human action than mere bodily movement: but dualists wrongly claim, and behaviourists rightly deny, that that 'more' consists in a hidden and separate mental ingredient.

But how can we accept *both* the dualist claim that there is 'more' to human action than mere bodily movement, *and* the behaviourist claim that that 'more' does not consist in a hidden mental ingredient? Only by rejecting their shared assumption that observation, knowledge and philosophical analysis must begin with physical bodies and movements: that these are what we directly observe; that these provide the basic data from which we must construct our knowledge of other people and their actions and intentions; that these must be basic elements in a philosophical analysis of the concepts of person and action. We must claim instead that we *begin* with people and their actions: that these are what we can directly observe, and directly know; that these are not reducible by philosophical analysis

to such supposedly simpler or more basic constituents as bodies and their colourless movements.

As I sit in the pub, I see and hear other people engaged in their various activities; buying drinks, chatting, playing darts. I see Ian having an argument with Pat, and see him punch her in the face; this is what I tell the police when they ask me what I saw, and the court when I appear as a witness at Ian's trial. On the dualist view, these descriptions of what I observed report the multiple inferences which I made from the physical bodies, movements and sounds which, strictly speaking, I directly observed: but we should rather say that they report what I directly observed; other people and their actions. I saw Ian assault Pat 'with my own eyes'; I heard their argument with my own ears: I did not need to infer an intention to hit Pat from the bodily movements which I observed, and thus infer that this was an assault; I saw the assault, Ian's intentional action, itself.

My description ascribes to Ian an intention to hit Pat, but does not refer to something happening in the hidden realm of his mind. I do not need to know what passed through his mind as or before he hit her, since I see his intention *in action*; the intention is identical with, not something separate from, his observable action of hitting Pat. But this is not to say, with the behaviourist, that in ascribing that intention to him I am simply describing or predicting a pattern of colourless bodily movements: if I saw only such movements, I could not ascribe an intentional action to him at all; but what I see is an intentional action.

To remind ourselves in this way that we do typically claim to see and hear people and their actions is to remind ourselves not just, as the dualist must argue, of the mistaken *beliefs* which we hold about what can be directly observed, but of the *meanings* of the concepts of person, action and intention: for the meanings of those concepts are given in their ordinary usage; and ordinary usage shows that persons and intentional actions are directly observable. Persons and actions, that is, are logically *basic* categories; these concepts cannot be explained by an analysis which seeks to reduce them to supposedly simpler elements.

There is, of course, a sense in which persons or actions consist of bodily and mental aspects: we can describe the physical or the mental aspects of a person (she weighs seven stone; he thinks of Jeannie) or an action (his arm moved; she intends to hit that target). But to do this is not to isolate distinct ingredients which make up a person or an action; rather, it is to abstract certain aspects of the *unitary* concept of a person as an embodied thinking being, or of an action as an intentional engagement with the world. The concepts of person and action are not constructed out of some more basic notions which are given to us in experience: we do not *begin* with the

concept of a body or bodily movement, and then *add* further ingredients to reach that of a person or action. Our experience and observations are structured by such concepts as those of person and action: to see persons merely as bodies, or actions merely as bodily movements, involves a difficult process of abstraction from what we initially see and know, not one of analysing out the simpler ingredients of a complex whole.

This anti-dualist (and anti-behaviourist) view clearly requires more explanation than I can provide here:⁸ but it can be clarified by looking again at the features of our ordinary thought which seemed to favour a dualist view.

First, there is, of course, often room for doubt or mistake about a person's intentions, even in simple cases like that described above. I am not sure whether Jane is moving towards the combatants with the intention of stopping or of joining in the fight: even my belief that Ian intended to hit Pat could be mistaken; perhaps they were talking about the title fight, and he hit her accidentally in demonstrating the champion's left hook. When there is room for doubt about what an agent intends, we may have to infer his intentions from the available evidence – from 'what he did, what he said, and all the circumstances of the case'; and our inferences may be mistaken. But such inferences are neither *from* colourless bodily movements, nor *to* the contents of a hidden mental realm to which only the agent has direct access: they are from the actions, or aspects of actions, which we observe, to the broader patterns of meaning of which they are part.

To discern an agent's intentions is to grasp the relation between her action and its context (including what else she does); what she will count as success or failure in what she does; and the truth of a range of hypotheticals about what she would do if . . . ; and we may be uncertain or mistaken about her intentions in so far as we are ignorant of or mistaken about any of these matters.

Suppose we know that Mrs Hyam set light to petrol which she had poured through Mrs Booth's letterbox, but are as yet unsure what her intentions were in doing that (note that we begin with knowledge of her intentional action of setting light to the petrol, not merely of her colourless movements; had we been there we would have *seen* her set light to the petrol). To discern her further intentions in acting thus, we must grasp the context of that action (her relations with Mrs Booth and Mr Jones); the

8 See J. Cook, 'Human beings'; L. Reinhardt, 'Wittgenstein and Strawson on other minds'; A.I. Melden, *Free Action*; I. Dilman, *Matter and Mind*, Part II; and pp. 158–63, 201–4 below.

broader pattern of actions of which it is a part (her journey to and from Mrs Booth's house; her precautions against being detected, her failure to alert the fire-brigade); what she will count as the success or failure of the action (that she will count it as a success only if it sets fire to the house and frightens Mrs Booth); what she would do if, for instance, she found that Mrs Booth had moved away; how she would react if, for instance, the house did not catch fire. We discover her intentions by locating this particular action within a broader pattern of actions and reactions; by relating it to an end (frightening Mrs Booth into leaving town), and by relating that to its own wider context.

That wider context includes, of course, her own beliefs, desires and responses. But these are themselves shown, or could in principle be discerned, in her actions: in what she does, in what she says (or would say), and in how she responds or would respond to what happens. We may of course get things wrong, either because we misinterpret her actions, or because she deceives us. But our mistakes do not concern the contents of a mental realm which is, in principle, always hidden from us; they concern what could in principle be adequately known, if we knew more about her actions and their context. For we are mistaken or deceived about the *meaning* of her actions; and that meaning is, in principle, discernible in the larger pattern of her actions and her responses: we may not *in fact* be able to discern it, but this is not because it is *necessarily* hidden from us in a separate mental realm.

Dualism portrays the interpretation of human actions on the model of the scientific explanation of empirical phenomena, as a matter of discovering the unobserved causes of observed effects. A better model would be the interpretation of books or works of art. When I read a philosophical book, what I *see* are not mere marks on paper, but words and sentences. In working out the book's meaning, I am not trying to make inferences from what I read to some separate realm of meaning: I am trying to identify the pattern and direction of thought which can be discerned *in* the book, given the wider context of thought in which it is set; and my account of its meaning will show how its parts are related to each other and to that wider context. So too, in trying to understand a person's actions (what he is doing and why), I am trying to see what they mean; to discern the pattern of which they are part, their relation to their context, and the direction in which they are moving.

That pattern may be manifest in what I can see, as when I see Ian hit Pat. Even here there is room for error, as we have seen; I might see as a deliberate blow what was in fact a pugilistic demonstration which misfired: but my error then consists in misreading the action and its role in their

discussion; and I correct it, not by learning what was happening in the hidden realm of Ian's mind, but by gaining a better understanding of the action's context and of its character in that context. In other cases we must infer an action's meaning (an agent's intentions) from more limited evidence: we do not have all the pages of the book, and must reconstruct its meaning from what we have. Sometimes this is quite easy: if we know that a person waited on a bridge with a block of concrete, and pushed it off the bridge when he saw a car about to pass underneath, we could ask 'what else could a person who pushed such objects have intended but to cause really serious bodily harm to the occupants of the car?' (*Hancock and Shankland*, p. 469); given what we already know of the action and its context, we can readily discern the end towards which it is directed. Sometimes the task is more difficult, or even impossible, if we know little of 'what he did, what he said, and all the circumstances of the case' – as when we try to reconstruct a book and its meaning from only a few pages, and without any full knowledge of its context: but the character of the task remains the same; it is that of finding the meaning which is, albeit incompletely, manifest *in* his actions.

Second, we must of course explain not only intentions as they are revealed in actions, but bare intentions which have not yet been put (and may never be put) into action. I have argued that we cannot portray bare intentions as inner mental states which are logically distinct from action: for intention is logically parasitic on action; it is necessarily directed towards action, and can be understood only in terms of its relation to action. A thought of the form 'I will do X' amounts to the expression of an intention (and not merely an idle thought about the action) not in virtue of its intrinsic character as a mental occurrence, but only in virtue of the way in which it is related to the actual doing of X.

We can compare bare intentions to promises, as ways of *committing* myself to an action. A declared intention to do something ('I intend to mark your essay by tomorrow') may indeed be intended and taken as a promise to do it (you will rightly complain if I fail to mark your essay by tomorrow): but even when I do not announce my bare intention to others, or so qualify it ('I intend to do it, but . . .') that it does not amount to a promise, to form a bare intention to do X is still to commit myself (perhaps only qualifiedly) to doing X; hence the fact that if I do not carry out that intention I may be criticized for, or at least be asked to explain, my failure to do what I intended to do. To portray bare intentions as commitments is to emphasize the central point that their meaning consists, not in their intrinsic character as mental occurrences, but in the way in which they relate an agent to a future action: a bare intention is a bond by

which I tie myself to a future action, or the shadow which that future action casts into its past (see M.H. Robins, *Promising, Intending, and Moral Autonomy*).

Bare intentions, and intentional actions, of course often involve various kinds of thought about the intended action: I may think about what I am going to do, and about what I am doing as I do it; and some of my thoughts may indeed express my intentions. But it remains true that no particular thought (or feeling or other mental occurrence) can amount to an intention in or by itself: just as particular acts (raising my arm) have their character and meaning as actions only in virtue of their role within a wider structure of action and context (signalling to turn left, or waving to a friend), so my thoughts take their character and meaning as expressions of intention from their relationship to the action which I intend or in which I am engaged.

Third, we must also explain an agent's authoritative knowledge of her own intentions. I surely do have an *immediate* knowledge of what I intend to do – both of my bare intentions and of the intentions with which I now act; a knowledge which does not depend, as an observer's knowledge must depend, on observing and interpreting my own conduct (I shall leave aside here those problematic cases in which we ascribe to an agent an intention of which she is unconscious, or about which she is deceiving herself). The dualist explains this knowledge as a matter of my privileged access to a mental realm which is necessarily hidden from others; how can it be explained if we reject Dualism?

It can be explained by drawing a sharper distinction between the kind of knowledge which I have as an *agent* and the kind of knowledge which I can have as an *observer*. The dualist portrays my knowledge of my own intentions, like my knowledge of the intentions of others, as being based on *observation*: the difference between them is just that I can directly observe my own mind, whereas I can observe only the external behaviour of others. We should rather say, however, that I know my own intentions not as an *observer* of my own mind, but as the *agent* whose intentions they are: I know my intentions *in* forming and acting on them. There are not two distinct processes involved, one of forming intentions and acting on them, and another of observing those intentions and thus coming to know what they are: rather, to form and act on an intention is itself to know, as an agent, what I intend. An agent's authoritative knowledge of her own intentions thus has to do, not with her privileged status as a direct observer of her own mind, but with her privileged status as the agent of those intentions and of the actions which are structured by them.

I have in this chapter offered only a bare sketch of some of the objections which, I think, undermine Dualism; and of an alternative view which

insists (against both Dualism and Behaviourism) that what we directly observe are not bodies and their colourless movements, but people and their actions. Both the objections to Dualism and the alternative view require far more discussion than I can offer here: but I hope at least to have cast some doubt on the dualist view which many jurists presuppose, and to have pointed the way towards a better way of understanding intention and its relation to action.

The issues discussed in this chapter do not impinge directly on the questions about the meaning of intention, and about its role as the key determinant of criminal liability, with which the last three chapters were concerned, since the answers which I have suggested to those questions do not depend on the arguments against Dualism which I have offered in this chapter. These issues do bear on the question of how intention is to be proved; and they are also relevant to the orthodox distinction between *actus reus* and *mens rea*, in so far as that distinction reflects dualist assumptions: but these are not matters which we can discuss here (though the discussion of recklessness and of criminal attempts in the next two chapters will depend in part on the argument that we should not draw as sharp a distinction between '*mens*' and '*actus*' as many jurisprudents draw). We must instead move on to consider some of the ways in which criminal liability may be extended beyond the paradigms of responsible agency with which we have so far been concerned.