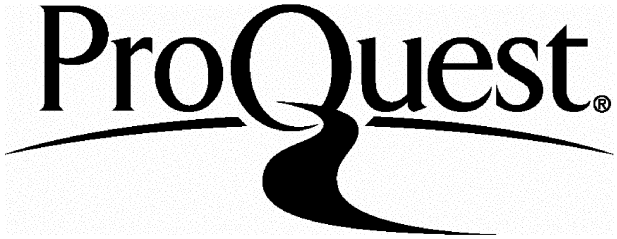
# AGENCY AND THE CAUSES OF ACTIONS

Zoë Victoria Payne

University College London

MPhil. in Philosophy, 1999



## ABSTRACT

This thesis attempts to outline a causal theory of human action. This theory is intended to be a partial explanation of what action is, and, specifically to illuminate the relation between actions and reasons. The starting point is Davidson's theory of action. Problems with this theory will be identified and distinguished. A set of conditions that a satisfactory theory of action must meet will be specified. First, that the theory should preserve the intuition that the relationship between reasons and actions is in some way causal. Second, that the theory should explain what actions are. Third, that a directness constraint be specified, that eliminates dependence on causal chains, and hence the possibility of deviant causal chains. Fourth, that the theory give a plausible rôle for the agent in his actions. Fifth, that the theory should not provide definite answers to questions of whether events are indeed actions, or indeed intentional actions, where our ordinary talk of actions and intentional actions does not provide such answers. I will consider what actions are, given that they cannot be whatever reasons cause. Hornsby's discussion in her *Actions*, will provide the focus of the debate. Pursuing the investigation of Hornsby's ideas, the possibility that the agent could be introduced into action by the device of describing actions as tryings will be considered. It will be concluded, first that actions or tryings are not the causes of bodily movements, and second, that a description of actions as tryings is not appropriate for the formulation of a causal theory of action. It will be argued that it is not strictly true that reasons cause actions, but rather that actions occur when reasons directly cause bodily movements, where the reasons and bodily movements are of a sort explained. An account of how a causal theory of action might involve the agent, based on Velleman's ideas, will be included in the theory. Finally, it will be argued that the final theory explains the possibility of uncertainty of whether an event is an action or intentional action.

# CONTENTS

## ACKNOWLEDGEMENTS

I am extremely grateful to my supervisor, Dr Lucy O'Brien, for her inspiring comments and discussions of many of the topics covered in this thesis. I would also like to thank all the other faculty members and my fellow students at University College London whose interesting and useful comments on various earlier drafts of parts of this work have led, I hope, to significant improvement and clarification of my thesis. Thanks also to Professor Jennifer Hornsby, of Birkbeck College London, for letting me have a manuscript of her forthcoming paper 'The Poverty of Action Theory'.

Finally, I gratefully acknowledge the Humanities Research Board of the British Academy, whose funding allowed me to undertake my MPhil. and so to write this thesis.

# Chapter 1. Introduction

The purpose of this thesis is to answer David Velleman's question, 'What happens when someone acts?' (Velleman, 1992a). Like Velleman, I expect the answer to give a causal story, relating reasons, actions, physical events and the agent himself. Not all these things need be identified with events in the causal sequence, but it should be clear how they are related, and, if they are not events in the causal sequence, then how the operation of that sequence amounts to their occurrence or involvement.

## *1.1. A theory of action*

It is my aim to propose a theory of action, as characterised by Hornsby (forthcoming). Action theory in the analytic tradition over the past three decades or so is, she says, characterised by two features:

> First: it treats actions as particulars; these particulars are taken to be a species of events, of once-off occurrences. Secondly, action theory styles itself as *theory*, because it makes generalizations about actions, and about the relations that actions stand in particulars of other sorts. (Hornsby, forthcoming)

The work presented here is very much in this tradition, not least because it takes as its starting point Davidson's account of actions and reasons in 'Actions, Reasons, and Causes' (Davidson, 1963). Davidson's view will be discussed in Chapter 2, as will problems with his account that have been identified by various thinkers. I will attempt to identify a strategy that will allow modification of Davidson's account to avoid these criticisms. To this end, I will specify five conditions that a plausible theory of action must satisfy. Condition (1) will simply specify that the theory should be causal, by which I mean that the relations between particulars mentioned should be causal. I hope to avoid the conclusion reached by some thinkers (for example, Kim), that the relationship between reasons and actions is of some special kind. Further, I wish to distinguish my project here from

a conceptual investigation of action, and the various concepts related to it.

## *1.2. Acting for reasons*

In order to construct a theory of action (or anything else) it is necessary to have some rough ideas regarding the subject matter that is to be explained by the theory. In the case of action theory, these ideas should be provided by our everyday talk regarding action. Deciding what actions are on the basis of our usage of the word 'action', is not straightforward. We do not usually speak of people *acting* (unless they are in a play or pretending), but rather of people moving, running, turning on lights, going to the cinema, and so on.[1] It has been suggested that 'there is an action whenever someone does something *intentionally*' (Hornsby, forthcoming; Davidson and Anscombe have a similar view). I agree with this, but do not think that someone acts *only* when they do something intentionally. I do think that someone acts only when they do something for reasons, and I think that things done intentionally are a subset of things done for reasons. For example, suppose I am giving up smoking, and am sitting at a table on which there is a packet of cigarettes (perhaps of the same brand I used to smoke). Suppose, further, that I take a cigarette from the packet and light it (with a lighter that was also on the table). It seems possible that I might do this without being aware of what I am doing, and that I suddenly realise with some surprise that I am holding a lit cigarette. In this case, I do not seem to light the cigarette intentionally, nor need I have been doing anything else intentionally. I suppose it might instead be the case that my lighting the cigarette was reflexive, something that did not require any explanation in terms of intentional (in the sense of having content) states of the agent, in which case we would not want to call it an action. However, I do not think such an account would be plausible in most cases of the sort I have described.

[1] This point is made by Hornsby (1980).

It is easy to imagine that, say, a knee-jerk, or a flinch, is a reflex, but not more complex behaviour like taking a cigarette from a packet and lighting it. I think a better explanation in most cases would be that I lit the cigarette for a reason, (perhaps I wanted a cigarette—consciously or unconsciously, or I performed the action from habit[2]). If it transpired that there was no reason for my behaviour, I would then be disinclined to call it an action.

Thus, in my view, actions are things done for reasons. This is why I take the relationship between reasons and actions to be one of the most important areas for investigation in the philosophy of action. It is also why Davidson's theory that reasons are the causes of actions is an excellent starting point for the discussion. However, if reasons are the causes of actions, and actions are things done for reasons, then, unless we implausibly want to claim that whatever a reason causes is an action, it seems that we have some work to do to escape circularity. This is the motivation for condition (2): the theory must say what actions are, i.e. just what sort of 'thing' an action is. Then we might be able to say that, when such a thing is caused by a reason, then it is an action. In fact, I will argue that we cannot settle what an action is in this way, but rather that what an action is should be explained by the completed theory. Actions will not be elements in this theory, but the discussion of what actions are, and how they relate to bodily movements, to be found in Chapter 3, will make an important contribution to the final theory. I will argue that reasons do not cause actions, but to cause certain sorts of (mainly) bodily movements, and further, that they should directly cause these bodily movements. Condition (3), therefore, will specify that a definition of a reason's *directly* causing a sort of bodily movement must be given, though this condition will not be met until the final chapter, in §5.2.

[2] According to Velleman (1989, ch. 7b), habits can be reasons.

## *1.3. Acting intentionally*

The difference between acting unintentionally for reasons and acting intentionally for reasons seems consist in my awareness of what is going on: when I act intentionally, I know my reasons for so acting, and act for these reasons I know I have. I will not directly address the problem of when an agent acts intentionally, but will subsume it into the question of how an agent is involved in his action. When he acts intentionally, it seems, he must be involved to the extent that he is aware of his acting, and he knows his reasons for so acting. I will also occasionally mention the idea of 'full-blooded' action. Some writers have taken this to simply mean intentional action, but I will take it to mean more: that the action was premeditated and performed as a result of a prior intention. Thus, an agent's involvement in his action might go further than simply his awareness of performing it and his knowing the reasons for his so acting. Condition (4) demands that a plausible account of an agent's rôle in his action be given, and I take it that meeting this condition will provide us with the resources to say when an agent acts intentionally.

I will address this condition first in Chapter 4, where I will discuss views of O'Shaughnessy and Hornsby, which suggest that actions are partly or wholly tryings. This sort of approach seems a plausible way of involving the agent in his actions in the right way: if I try to do something, then surely I know what it is that I am trying to do, I am aware of my action and of the reasons for so acting. However, I will argue that there are problems with both O'Shaughnessy's and Hornsby's accounts, and that describing actions as tryings, unfortunately, is not compatible with my aim of outlining a causal theory of action of the sort I have specified. Instead, I will adopt a variation on Velleman's position (1989, 1992a). This will be discussed in the final chapter.

## 1.4. *A general theory of action*

I want the theory of action I propose to explain what it is to do something for a reason, that is to act, without entailing that whenever an agents acts, he does so intentionally. However, I also want it to explain how an agent can be involved in the action, and thereby to act intentionally. I take 'action' to apply in a wider set of cases than many of the philosophers who have worked on the philosophy of action. My project is quite different from that undertaken by, for example, Thomas Pink (1996, 1997). His aim is to provide an account of distinctively human action (an aim shared to some extent by Velleman). Pink compares his view with Aquinas', and against Hobbes'. He appears to think that there is a class of actions that only humans are capable of, and, further, that this sort of action is radically discontinuous with simple animal action and requires a very different sort of explanation. On this point, I am with Hobbes and against Aquinas and Pink. Though I do agree that there is a roughly defined class of actions which we recognise only humans as being able to perform (actions that require complex planning, prior decisions and intentions, for example), I find it hard to see them as radically different in kind to common-or-garden animal action (as performed by animals that can act, or humans when we are not being distinctively human). Rather, I see this subset of actions as having the same basic features as animal action, but with further features as well. My hope is that a theory of action can be formulated which accounts for actions that do not share these features, but that can be extended to cover the additional features which actions may have and which we feel are important. It is outside the scope of this work to discuss decisions, and therefore Pink's work, so I will only be able to go a short way to showing how the basic theory could be refined to explain the fact that we can make decisions on how to act long before we actually do, but I do believe that such refinement is possible. I will often say that I am seeking a 'general' theory of action, and I mean by this that I want a theory of action that explains all action, but can be added to in order to explain

additional important features of some actions. I will argue for a theory that includes among actions all things that are done for reasons, and show how elements can be added to this theory to explain things that are done intentionally, or how agents might be involved in actions. Hornsby says:

> Most action theorists introduce assumptions and modes of speech which ensure that action theory's most widely accepted tenets could not be even a component of any properly rich account of ourselves and our place in the world. (Hornsby, forthcoming)

When I say that my aim is to outline a general theory of action, I also mean that I seek a theory which avoids this mistake. While I do not believe for a moment that the theory I outline tells the whole story about action, I want it to at least tell a part, and not in such a way that it is a barrier to the whole.

The final condition that I hope to satisfy is linked to the idea that there is no special, radically discontinuous, class of actions that is distinctively human and of which a theory can be constructed whilst ignoring actions more generally. I take both action and agency to be matters of degree, and condition (5) will specify that this should be accounted for. For example, someone who believes that actions only occur when someone does something intentionally might disagree that my lighting the cigarette in the above example was an action, and while I think that it is, I am sympathetic to their doubts. The answer might be that my lighting the cigarette lies somewhere between a reflex and an action. I want to be able to give this answer to a critic of my, admittedly rather weak, notion of action. I will also argue that whether an agent is involved in his action, or how far an action is intentional, is something that we sometimes have no definite answer for in ordinary language. We might say that the agent was involved in this way, but not in another, or that the action was intentional insofar as it had these features, but not intentional insofar as it lacked these others. Given that ordinary language explanations of action are all we start by knowing before we embark on the philosophy of

action, if the theory we propose were to give definite answers where ordinary language gives none, we would be forced to question whether we were indeed talking about the same things as are talked about in ordinary action language.

## *1.5. Conclusion*

My purpose, then is to outline a theory of action that meets the five conditions which I have outlined, and for whose justifiability I will argue. In brief, the strategy I will follow is, in Chapter 2, to set the context for the discussion, using ideas of Davidson's (1963) and Velleman's (1992a). The position at this point will be that reasons cause actions. I will then discuss some problems with this position, and identify the five conditions which I believe that, if satisfied, would remove the barriers to a plausible causal account of action. In Chapter 3, I will consider what actions are, in the hope that this will allow us to see whether it is possible that reasons cause them. While I will come to some conclusions about the individuation of actions, and how we might identify basic actions, I will not be able to say by the end of Chapter 3 what actions are. Chapter 4 will be devoted to the discussion of the idea that actions are, partly or wholly, tryings. This discussion will be motivated by the idea that trying seems to involve the agent. However, whilst I will agree that most actions are tryings, that describing them as such will not further the aim of presenting a general theory of action. Finally, in Chapter 5, I will pull together the conclusions of chapters 3 and 4, and discuss a different account, inspired by David Velleman, of how the agent might be involved in his actions. Chapter 5 will contain the outline the theory I propose, and I will also try to show there how it meets the five conditions I specified.

# Chapter 2. A causal theory of action

In this chapter, I will present the causal theory of action on which I want to build in this thesis. I will discuss Davidson's seminal paper, 'Actions, Reasons, and Causes' (1963), and briefly talk about the type of view of causation assumed by this paper. I will proceed to consider one form of objection to Davidson's theory, based on the possibility of deviant causal chains. I will argue that these sorts of objection require the causal theorist to provide some account of 'directly' caused action, but that this is not the whole solution. Following Velleman (1992), I will suggest that the deviant causal chain counterexamples to Davidson's theory bear some similarity to another set of problem cases, namely the examples of an agent's alienation from his action given by Frankfurt (1988). I will therefore conclude that deviant causal chain problems do indicate the sort of problem that Chisholm (1966) intended to highlight when he constructed one of the first examples. This is that causal theories like Davidson's do not present the agent as author of his actions and that, in this, they are at odds with the way we ordinarily speak of actions. I will then consider what strategies might succeed in introducing the agent into action.

## *2.1. The standard story of human action*

In his 1992 paper, David Velleman asks, 'What happens when someone acts?' and characterises a standard answer to this question as follows:

*2.1.1*

> There is something that the agent wants, and there is an action that he believes conducive to its attainment. His desire for the end, and his belief in the action as a means, justify taking the action, and they jointly cause an intention to take it, which in turn causes the corresponding movements of the agent's body. Provided that these causal processes take their normal course, the agent's movements consummate an action, and his motivating desire and belief constitute his reasons for acting. (Velleman, 1992a; p. 461)

This is what Velleman calls the 'standard story of human action' and it will provide the context of our debate. Donald Davidson's papers collected in 'Essays on Actions and Events' have been the inspiration for much of recent thought about action, though other thinkers, perhaps most notably G. E. M. Anscombe, have been responsible for some very important ideas in the field. Davidson argues for something like the standard causal story characterised by Velleman. Davidson's position in one of his first papers on action, 'Actions, Reasons, and Causes' (1963), did not quite fit the structure of the standard causal story as Velleman has it, as Davidson did not include intentions in his account. At that point, Davidson thought that intentions would be reducible to other psychological states, such as beliefs and desires, though he later changed his mind. In the next section I will examine Davidson's original theory, and then discuss ways in which this could be modified and the merits of so doing.

## *2.2. Actions, reasons and causes*

In 'Actions, Reasons, and Causes', Davidson introduces the term 'pro-attitude' for any motivating psychological state, such as desires, wants, values and so on. Unless directly quoting or discussing Davidson, I will instead use 'desire', but I mean it in the broad sense in which Davidson means 'pro-attitude' to be understood. Davidson starts from the consideration that when we act intentionally, we act for reasons. A reason for an action *A*, Davidson goes on to claim, is a belief-desire pair that together rationalise *A*:

*2.2.1* *R* is a primary reason why an agent performed 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. (Davidson, 1963; p. 5)

When an agent intentionally *A*s, he has some pro-attitude towards actions of a certain kind and a belief that *A* is of that kind. It could be that the agent has a pro-attitude towards actions that have the property of being arm-raisings, and trivially but nonetheless correctly, believe that

the action of raising his arm has this property. More substantially, the agent may have a pro-attitude towards actions that have the property of being signallings, and believe that the act of raising his arm in the current context would have that property. If *A* is some action described in a way that does not make it misleading to say an agent intentionally performed it, it seems plausible that he must have pro-attitudes and beliefs that rationalise it.

However, that an action *A* is rationalised by a belief-desire pair and action *A* is in fact performed does not entail that the primary reason comprised by that belief-desire pair is the reason for that action. Statement 2.2.1 gives a necessary but not a sufficient condition for something's being a reason for an action *A*. Our agent may want to lose weight, Davidson suggests, and believe that exercise would promote this end, *and* the agent might also exercise. However, it does not follow that the agent exercised *because* he wanted to lose weight: his exercise could have been motivated by a quite different pro-attitude, such as the desire to work off frustration, or the sheer love of exercise. That is, it is possible to have a primary reason *R* for performing an action *A*, and perform that action, and yet not perform it for *R*. In short, that I have reason *R and* do *A* does not entail that I do *A because* of reason *R*.

Davidson goes on to suggest that the link between reasons and actions is causal. He says:

*2.2.2* A primary reason for an action is its cause. (Davidson, 1973; p. 12)

After all, we say that we act *because* of certain reasons, and often when we say "*q* because *p*", we mean that *p* caused *q*. For example 'the window broke because the stone hit it'. Accordingly, perhaps 'I raised my arm because I wanted to signal, and I believed that raising my arm would accomplish it' is a causal explanation in the same way: that is, that my pro-attitude and my belief cause my action. A further indication that causation might be the correct relationship between reasons and actions is that we often think of causes as necessary and sufficient for their

effects, and this sort of thinking seems to apply fairly well to reasons and actions. If I want to signal, and believe that raising my arm will do it, have no other reason for raising my arm, and as a result raise my arm then it seems true that, first, if I had not wanted to signal (*ceteris paribus*), or had not believed that raising my arm would constitute signalling, then I would not have raised my arm: the reason was necessary for the action. Second, all that was required in the circumstances for my raising my arm was my wanting to raise my arm enough to raise it: the reason was sufficient for the action.

I am not suggesting here some sort of simple reason determinism after the manner of Churchland (1970), that is, that if we attach more weight to one reason than any other we will act on it. All I am suggesting is that in any particular situation in which we act for a reason (and do not have other reasons for the same action that would have led us to act), we normally think that without the reason we would not have acted, and nothing else other than the reason was needed in order for us to act. We normally think that we could, for example, raise our arm if we wanted to. 'Particular situation' covers a multitude of other possible facts: perhaps that I did not have any stronger reasons for alternative actions, that my body and mind were in such a state as to allow me to perform the action, among others. Nonetheless, I recognise that my claim that reasons are sufficient for action might be controversial. One objection might be that it is not enough to have a reason, we also have to decide to act upon it, or form an intention to act upon it. I am trying to leave discussion of intention until later in this chapter, but will say now that this does not threaten my contention: presumably all the factors that lead to my forming the intention are there in the 'particular situation'. What would threaten my contention that reasons are sufficient for actions is if it is possible to be in exactly the same psychological state, in exactly the same physical environment, and yet act, or decide to act, differently. I am not convinced that this sort of radical freedom is either supposed by our ordinary discourse, or is the case, but, even if it is, this does not rule out a

causal theory of action, only a deterministic causal theory of action. My argument here is simply that there are *prima facie* similarities between ordinary reason explanation and causal explanation. Whether this similarity is sufficient for the construction if a causal theory of action can be best judged by considering the plausibility of a causal theory of action, which is what I propose to do in this thesis.

The nature of causation and causal explanation is a large and important philosophical topic in its own right. This is one reason why the question of whether reasons are causes cannot be simply answered. Nonetheless, it will be necessary here to have some view on the nature of causes and causation, though it is possible that a wrong choice of view could undermine substantial parts of my argument. The influence of Davidson in the philosophy of action makes it natural to pick his view of causation as a relation between particular events. This view of causation is adopted by many philosophers in the field of action, including Jennifer Hornsby (1980) whose work I will be discussing at some length. A radically different view of the nature of causation would make it difficult to engage with their arguments.

## *2.3. Can primary reasons be causes?*

However, I do find it hard to make sense of a view of causation which mentions, in any one case in which causation happens, only two events: the cause and the effect. Jennifer Hornsby (Hornsby, 1993; pp. 288-289) has registered her doubt as to the intelligibility of saying that a primary reason constituted by a belief-desire pair is the cause of an action. She says (p. 288):

> If we are to make sense of it as 'the cause', then the primary reason must be an item that there is if and only if the relevant agent believes some particular thing and desires some other (related) thing. But why should we think that there is any such item as this? Why should the acknowledgement that we say something about what she believed and desired in causally explaining why she did what she did lead us to accept the existence of anything that 'the cause of her action' stands for?

I take Hornsby to mean that we have no reason for thinking that 'the cause of her action' stands for some composite item, and in this she seems correct. However, there does seem to be an intuitively plausible way in which we can understand an event being caused by a belief and a desire. For example, if the cause of my concussion is that I stood up and at the same time you turned round holding a plank, my standing up and your turning round with the plank do not constitute a composite item that is the cause of my concussion. To what exactly 'the cause' of an event refers to seems to differ with context. First, it might refer to a situation and a change in that situation, such as 'the fire was caused by water dripping on faulty wiring'. Second, 'the cause' might refer only to the change, as in 'the window broke because the stone hit it' (we wouldn't normally include the fragility of glass in our description of 'the cause', though it is possible to imagine a situation in which we would, for instance, if the person to whom we were giving our explanation asked why the wall didn't break when the stone hit it, but the window did). Therefore, when we say that the primary reason for an action is its cause, we must be careful about what we mean. We cannot be suggesting that a belief-desire pair constitutes some item that is a cause, but must mean one of a number of less controversial things. The items we mention when giving the cause of an event can be characterised as dispositional or active. 'The cause' can refer to either the dispositional or active element, or the conjunction of both. Further, as in the case where I stood up as you turned round holding the plank, there can be more than one active element. It seems necessary, however, to have some change, active cause or event, even if it is not mentioned in a particular explanatory context. So when one says that the cause of an action is its primary reason, one might mean:

(a) I believe that action *A* has a certain property, I come to have a pro-attitude towards actions with that property, and the pro-attitude actively causes *A*, with the belief being a dispositional cause of *A*.

(b) I have a pro-attitude to actions with a certain property, but do not know what actions have that property. I come to believe that action

*A* has that property, and the belief actively causes *A*, with the desire being a dispositional cause.

(c) I simultaneously come to desire actions with a certain property and believe that action *A* has that property, and the two actively cause the action.

Davidson, as far as I understand, would argue that this does not pose any problem for his account, and that his reply would rely on the distinction between the items that are appropriately referred to in giving causal laws versus causal explanations. A full discussion of this matter is outside the scope of this thesis. I merely wish to establish that I take causation to be a relation between events, against a background of other events, and that it is a feature of causal explanation that we describe situations, or facts, that provide context for the causal relation between events. This would not pass muster as an account of causation, but I hope it is not inconceivable that some plausible account of causation could explain the features I have highlighted.

A consequence of this view is that while the primary reason for an action may be its cause, it need not be its whole cause. There could be a number of dispositional causes, or dispositional elements of the cause (depending on how we're counting causes), and even other active causes, or factors contributing to the cause, of the action. Later in this thesis, I will argue that there must be such additional elements in action. In order to be properly Davidsonian, these additional elements of the cause would have to feature in the causal story as events, rather than facts. I will attempt not to propose any causal relations that cannot be characterised as relations between events, though for convenience I will sometimes talk as though facts can be causes.

## *2.4. Deviant causal chains*

I have indicated how Davidson's account might be intelligible, but will now turn my attention to the question of whether it is plausible. One

common type of objection to a theory that claims reasons are the causes of actions was first highlighted by Roderick Chisholm (1966). A variant on his problem case is as follows:

*2.4.1* A man is driving home. He desperately wants to become rich, and it occurs to him that killing his wealthy uncle would have this consequence. He is so horrified at his thought that he loses concentration and runs over and kills a pedestrian. It transpires that this pedestrian is his uncle.

Here, the agent has a pro-attitude, to get rich quick, and belief, that murdering his uncle would accomplish this, that together constitute a primary reason for his killing his uncle. It happens that he does kill his uncle and is caused to do so by this primary reason, but his killing his uncle is nonetheless not done for the primary reason. Davidson later constructed an example with similar features, which is worth describing, both to allow variety in the discussion and as the two examples have subtle differences which may demand different treatment. Davidson imagines a case as follows:

*2.4.2* 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 as to cause 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. (Davidson, 1973; p. 79)

Like Chisholm's example, Davidson's makes the point that the primary reason's rationalising an action, and its causing that action, is not enough to guarantee that the action was performed for that primary reason.

## *2.5. Actions and bodily movements*

It has been objected that the deviant causal chain example given at 2.4.2 does not give a genuine example of an action being caused by a primary reason, only of a bodily movement being caused by a primary

reason.[3] This emphasises a distinction I have thus far let pass unremarked between the standard story of human action characterised by Velleman (see 2.1.1, p. 12 above) and Davidson's story. Velleman has reasons causing 'corresponding movements of the agent's body', whereas Davidson has them causing actions. I wish to defer most of the discussion of the significance of this difference until Chapter 3, but will make some preliminary remarks now. First, we must have a reasonably good idea of what events count as actions before we can formulate a theory of action but it seems less clear that a theory of action should have the concept of 'action' as fundamental. This, though, would be the consequence of simply specifying that the events caused by reasons must be actions, not mere bodily movements, in an attempt to escape the deviant causal chain counterexamples. Rather than this, one would hope that the theory of action itself would give us some criteria for an event's being an action. Second, a popular way of distinguishing actions from non-actions, inspired by Anscombe and Davidson, is that 'there is an action whenever someone does something *intentionally*' (Hornsby, forthcoming). But if, as I suggested in Chapter 1, the things that people do intentionally are (perhaps a subset of) the things they do for reasons, then it seems we must have an answer to the question of what it is to do things for reasons before we can decide (from the point of view of our theory) when there are actions. I do not, at this point, wish to rule out a reply to deviant causal chain examples which plays on the distinction between actions and bodily movements, and will consider variants on this reply in Chapter 3 and Chapter 4. Any such reply would, though, have to give a plausible account of what an action is, independently of its relation to reasons. As a result, I will resist any *definition* of action in terms of things done intentionally,

[3] Hornsby, in her forthcoming paper 'The Poverty of Action Theory' suggests that even the phrase 'bodily movement' suggests that someone has moved their body, which is not the case in the deviant causal chain examples given above. The objection may be made clearer by saying that there are no actions in the examples given by Chisholm and Davidson, only bodies moving. Whilst I have some sympathy with her contention, I do think the word 'movement' is usually free from any imputation of agency. We might, for example, say, 'The equations describe planetary movement'. I will, therefore, stick with the locution 'bodily movement', though I do mean it to refer to bodies' movings, not people's movings of their bodies.

and would even if, as a matter of fact, whenever someone did something intentionally there was an action.

## *2.6. Internal and external causal chains*

It has been argued that an account that specifies that the event caused by the reason must be an action is not invulnerable to deviant causal chain counterexamples. Moya (1990, p. 122) suggests that we adapt Chisholm's example. Instead of accidentally running over the pedestrian, we are to assume that the agent notices him, but, in order to get to his uncle's home and murder him, the agent realises he cannot spare the time to wait for the pedestrian to cross. He therefore runs over the pedestrian on purpose, but still does not intentionally kill his uncle. However, this sort of deviant causal chain does not seem nearly so damaging. Davidson talks about external and internal chains, and is quite dismissive of the former, calling them 'quaint' (1973, p.79). I am not sure that the action/bodily movement distinction works in the nephew example, for reasons I will mention shortly, and so question the motivation for this amendment to Chisholm's example.

I think an alternative example would be a better case for discussion. Davidson mentions an example of Daniel Bennett's. A man tries to kill someone by shooting him. His shot misses, but startles a herd of wild pigs, which stampedes and tramples the intended victim to death. This shares with Moya's amendment of the nephew case the fact that an intentional action (the shooting) takes place, and causes an intended effect (the killing). In Bennett's example, it is accidental that the action has its intended effects, and in Moya's that the action not only has its intended effects, but also accidentally brings about an effect that the agent had another reason for bringing about. These cases are quite different from Davidson's climber example, where the deviancy takes place before any action and has the result that there is no action, where

there is a 'nonstandard or lunatic *internal* causal chain' (Davidson, 1973; p. 79). I find quite plausible the idea, proposed by Armstrong (1973) and endorsed by Davidson, among others, that we should deal with external causal chains by insisting that they must reflect, at least roughly, the pattern of practical reasoning. In order to give a more causal account, I would suggest that, when he acts, an agent has a number of beliefs about the way his action will progress. For example, that the pulling of the trigger will result in the expulsion of a bullet in the direction of the target, that the bullet will hit the target, that the bullet hitting the target will kill him. These beliefs could perhaps have a causal rôle in the action, as a basis for the belief that composes the primary reason (perhaps, in this case, 'pulling the trigger will result in the death of the target). If some or all of these beliefs turn out to be false, it will affect what we can say the agent did intentionally. Once an effect unexpected by the agent occurs, then perhaps he cannot be said to do intentionally anything that is caused by this unexpected effect. This may be too brief a treatment, but I do feel that the problem of internal deviancy is by far the greater, and thus propose to concentrate on this from now on.

In Davidson's example, it is clear that there is no action, but we might be less sure of this in Chisholm's nephew example. I have said that, whenever someone does something intentionally, there is an action, and in Chisholm's example the agent is, presumably intentionally, driving. Driving is perhaps not a paradigm case of physical action: I can be driving even when sitting perfectly still. The question might therefore arise whether we are always acting when we are driving. I think that the answer must be a qualified "yes", though anything resembling a full consideration of this question is outside the scope of this work. If I am right, then the nephew's running over of his uncle is an action, albeit an unintentional one. This treatment of Chisholm's example makes it more complex than Davidson's climber example. Insofar as the pedestrian's being hit is caused by the nephew's intentional action of driving, the deviancy is external, but the driving alone was not enough to cause the

pedestrian's being hit: the nephew's inattention, caused by his reason for killing his uncle, was also required, and the deviancy here is internal. I will return to the discussion of this aspect of Chisholm's example in §2.8.

## *2.7. Directness*

An objection that might spring to mind at this juncture is that, in the deviant causal chain examples given at 2.4.1 and 2.4.2 (p. 19 above), the reason causes the bodily movement (or action) in a very roundabout manner, *via* some attitude or response to the reason. Thus, perhaps, we require a constraint that specifies that the reason directly or immediately cause the action. It might be thought that this tactic runs the risk of ruling out legitimate examples of acting for reasons. Often, we have a reason for performing an action, we might for example have a reason for turning on a light, and this may cause our turning on the light, but *via* a desire or a belief that flipping the switch will result in turning on the light. But this does not show that a directness constraint will not help us to overcome the problems resulting from deviant causal chain counterexamples, only that we have more work to do in specifying the format of the primary reason for an action. I have said (§2.2.1, p. 13 above), roughly, that a primary reason comprises a desire for end *E* and a belief that action *A* will bring about *E*. The primary reason is a reason for doing *A*. But it now seems that, for example, my desire to see, plus my belief that turning on the light will enable me to see, whilst a reason for turning on the light, is not an appropriate format for primary reasons if we are to specify that they directly cause actions. I propose that the means-end belief that composes the primary reason should make reference to an action the agent can directly perform. I take it that one of the things that agents can directly do is move their bodies. Therefore, my preliminary suggestion for the appropriate format of the belief that composes part of the primary reason in the case of my turning on the light would be along the lines of 'Moving my arm and hand in such-and-

such a manner will result in the turning on of the light'.[4] In general, the action for which the reason is a justification, in the case of a primary reason, should be an action which the agent can directly perform. Elaboration of this point will have to wait until the discussion of basic actions in Chapter 3, and specifically in §3.3, but it initially seems that, if we can eliminate the notion of a causal *chain* from reason to action, then we have frustrated the possibility of *deviant* causal chains.

Before continuing, I will deal with two more possible objections that a directness constraint will rule out events that would be considered actions as being actions. Consider the case of an agent 'psyching himself up' to perform an action, to bungee jump off a bridge for example. He might repeat "I'm going to jump", "I'm going to jump" to himself, concentrating on his desire to jump. Concentrating on the desire in this way could put him in a frame of mind that causes him, or allows him to jump, which he may not have had the nerve to do otherwise. But it is not clear here that, when the time comes and the agent acts, the primary reason does not directly cause the action, though admittedly it also causes a reflection on itself that contributes to the cause of the action. The primary reason alone was not enough to cause the action, the reflection on the reason was also required. A clearer case might be made by modifying Davidson's climber example. Suppose the climber did want to drop his companion, and had decided to do so. However, he was aware that he would not have the nerve to do so, but realised that if he concentrated on his intention he would become so horrified that he would make involuntary movements and be unable to prevent himself losing his grip on the rope. He therefore concentrates on the intention, and events proceed as he expected. His dropping his companion, we would probably think, is an action, but it is not directly caused by the primary reason. I accept this, and in response, would say that here we have reached the

[4] Davidson convincingly argues (1971) that it does not matter that there is no non-trivial way to fill in the notion of moving one's body in 'such-and-such a manner'. Often all we can say is, for example, that one should move one's body in such a manner as to turn on the light, or flip the switch, or tie one's shoelaces.

limits of even the general theory of action I want to propose. If we think about what the agent actually, actively did, there is no physical movement. This case might be characterised as a mental action that has physical effects, and I think it is sufficiently different from the sorts of physical action I take as my subject matter for my theory to be excused from covering it.

## *2.8. Agency and actions*

Thus far, I have considered two sorts of responses to the deviant causal chain examples. First, that we could clearly specify what an action is, and then add the constraint that the reason cause an action, not merely a bodily movement. Second, we could formulate and apply some sort of directness constraint. However, Chisholm thought that his example showed something else, namely that the causal theory described by Davidson could not account for the agent as author of his actions. His contention was that the problem with the nephew case is that, whilst the agent's desires and beliefs cause an action, the agent himself was not the cause of the action. Like Davidson (1971), I am not sure that the relation between agents and actions is properly considered as causal, but there does seem to be something to Chisholm's point. Whether or not an account of the agent's rôle in action is required to avoid the implications of the nephew example, or whether a directness constraint would be enough, that the agent is missing from the action seems to be a valid criticism. Velleman (1992a) suggests that, insofar as deviant causal chains highlight this particular problem, they share an important feature with a different sort of example.

I will give two examples of this new type of objection to the standard story of human action. The first is from Harry Frankfurt (Frankfurt, 1988; p.18 & p. 21, discussed in Velleman, 1992). Frankfurt asks us to consider the case of a drug addict who is trying to break her habit. She does not want any drugs, but on the other hand she craves them because of the addiction. She might unwillingly, and perhaps even unwittingly,

form the intention to obtain drugs, and so go and do so. Frankfurt argues that this action is not fully intentional, but that there is something wrong here. The agent both intended not to obtain drugs, and on another level, intended to obtain drugs. This shows that, in the most full-blooded cases, there is more to action even then making a prior decision that is then followed through. However, I think that we would count the agent's action of obtaining the drugs as intentional, and even that, whilst she was in some way alienated (in Frankfurt's term) from her action, she was nonetheless the author of her action. This example is one that shows that an action's being intentional, and that an agent's involvement in his actions, might be a matter of degree. Velleman constructs a slightly different example:

*2.8.1* I meet with an old friend to resolve some minor difference, but find myself giving irritable replies to his comments and we part in anger. My later analysis of events suggests to me that I didn't want to resolve the difference, and that I was using this encounter to sever the friendship. In this case, beliefs and desires caused an intention, and the intention appropriate behaviour, in a non-deviant way, but subconsciously, without my participation (1992a, p. 464).

What we would say about this example is even less clear. Velleman thinks that there is no full-blooded action here, and in this he seems right. Neither does there seem to be an intentional action, but there is an action, and it does not seem necessary that the agent is completely alienated from it, as his later analysis revealed that the action furthered an end of his, though there is a stronger case here than in Frankfurt's example for saying the agent is not the author of his actions.

In the standard story of human action, an agent is of course involved as a bearer of the desires and beliefs that cause the action, and Velleman suggests that it might be said that '[c]omplaining that the agent's participation in his action isn't mentioned in the story is ... like complaining that a cake isn't listed in its own recipe.' (19921, p. 462). However, Velleman argues that this is unsatisfactory from the point of view of our ordinary concepts and talk of action. He says:

> In a full-blooded action, an intention is formed by the agent himself, not by his reasons for acting. Reasons affect his intention by influencing him to form it, but they thus affect his intention by affecting him first and the agent then moves his limbs in execution of his intention; his intention doesn't move his limbs by itself. The agent thus has two roles to play: he forms an intention under the influence of reasons for acting, and he produces behaviour pursuant to that intention. (1992a, p. 462)

My contention is that the problem of including the agent as author of his actions does not arise only for full-blooded actions, but that, even in examples like Frankfurt's, we need to account for the rôle the agent plays, even though this rôle is perhaps more limited than in the case of full-blooded action. I see the scenario that a general theory of action has to explain as follows: reasons can result in (i) movements that are not actions (as in Davidson's climber example), (ii) actions which are unintentional and in which the agent is not involved (as is possibly the case in Chisholm's nephew example), (iii) actions which are not accurately characterised as unintentional, but are not clearly intentional, and in which the agent is possibly involved, but not undeniably so (as in Velleman's friend example), (iv) actions which are intentional, but not unqualifiedly so, and of which the agent is author, but lacks full involvement (as in Frankfurt's example), and (v) paradigmatic, full-blooded actions, which are intentional and the agent is fully involved in. There could well be finer distinctions than these five. It seems from looking over these examples that the questions of whether an event is an action, and whether or not the agent is author of that action, are not wholly independent. Though it might be the case that there are actions of which the agent is not author, as is possibly the case in the Chisholm example and in cases of absentminded action, like tapping one's feet, our confidence that what has occurred is indeed an action grows as it becomes clearer that the agent was its author and, further, was involved in the alleged action to the fullest extent.

In constructing a general theory of action, we must make clear what it is for an agent to be 'author of' or 'involved in' his actions. One answer might be that he is the cause of his actions, though this cannot be enough if his involvement is to admit of the degrees I have argued exist. Nonetheless, this is the sort of answer that Chisholm proposed. Discussion of what has been called 'agent-causation' is outside the scope of this thesis: its flaws have been amply pointed out elsewhere (see, for example, Velleman, 1992a), and I have already said that I am not sure the relationship between agent and action is properly understood as causal. What is required instead is a causal story that, in the appropriate cases, can be seen to amount to the agent being the author of, and involved in, his actions in the right way. Velleman argues that we should look for psychological states or events that play the 'functional rôle' of the agent (1992a). This may prove to be a useful indication of where we should look for the agent in action, though strict functional equivalence is not required if we are not sure the agent causes his actions. I have left this latter unargued for, and propose to leave open the question of whether an agent does cause his actions, or whether his involvement is of a slightly different nature. Velleman, looking for psychological states that may play this functional rôle of the agent, pinpoints two places at which an agent may intervene: in deciding how to act, and in producing the behaviour that constitutes his acting. This latter is particularly interesting, as it seems to suggest the investigation of what an action is, that I have already indicated as required to avoid deviant causal chain problems, may also shed some light on the agent's involvement in those actions.

## *2.9. Intentions*

The reader has no doubt noticed that there is still one more distinction between Velleman's standard story of human action and Davidson's story that I have not yet addressed. Velleman's story includes intentions as the causes of actions, whereas Davidson's makes no mention of them. The

reason for this omission is that, as I said in my Introduction, I do not believe it necessary for action that we form a prior intention, and that intention therefore has no place at the most basic level of a general theory of action. This is not to say that accounts that include intentions or prior decisions do not have a valuable rôle to play, simply that they serve to illuminate a subset of cases of action. Nonetheless, I cannot ignore intentions and decisions entirely. The general theory that I sketch must be amenable to the adding of elements that account for the subset of actions that are done because of an intention, or done as the result of a prior decision. However, refining the general theory in this way would be too lengthy and complex a task, and cannot be undertaken in this thesis. Instead, I will just make two points.

First, I take it that Velleman's standard story of human action, where a reason causes an intention which in turn causes a bodily movement can tell part of the story in some cases. Here, the reason causes an action *via* an intention, and this might be thought to threaten the viability of a directness constraint on action. I think, though, that a primary reason can either directly cause an action *or* an intention. If the action is then performed for the reason, it must be directly caused either by the reason *or* by the intention. This would increase the complexity of the theory, but clarifying what it is for agents to act for reasons in cases where there is no intention assumed could, I hope, provide us most of the resources we would need to include intentions in this story.

Second, introducing intentions into the story does not help us address the problems raised in this chapter. The problem cases given by Frankfurt and Velleman already include intentions, and those of Chisholm and Davidson could easily be amended to allow for intentions. Shock at the *intention* to drop the rope or kill the uncle might trigger the causal chain that leads to the rope being dropped or the uncle killed.

I will not, therefore, investigate Velleman's first suggestion for the rôle an agent plays in action: that of forming 'an intention under the influence

of reasons for acting' (1992a, p. 462), though I accept that a full theory of action will account for this rôle as well as that of actually producing the behaviour he has reasons to produce.

## *2.10. Conclusion*

In this chapter, I have sketched the sort of causal theory which I will be attempting to refine in this thesis. I have also indicated the problems which the refined theory must avoid. We are now in a position to give the conditions that the sort of general theory of action I want to construct must satisfy. It must:

(1) Be causal.

(2) Give an account of what an action is.

(3) Give some sort of directness constraint.

(4) Give a plausible rôle for the agent.

(5) Admit of degrees.

My strategy will be to attempt to satisfy condition (2) first. To this end, I will consider the illuminating distinction made by Jennifer Hornsby (1980) between transitive and intransitive uses of the verb 'to move', the individuation of actions, and how basic actions should be identified. This discussion will indicate how we might satisfy conditions (3) and (4). It will also motivate the discussion of trying in Chapter 4, also inspired by Hornsby's (1980) work, which is a more explicit attempt to show how an agent may be involved in his actions. I will conclude, however, that identifying actions with tryings does not give us the resources to meet condition (5), and to some extent frustrates the effort to meet condition (1). In the final chapter I will attempt to use the ideas that have arisen from the discussion in Chapter 3 and Chapter 4, and add to them some ideas inspired by Velleman's (1989, 1992a) work, in order to sketch a theory that satisfies all five conditions.

# Chapter 3. Actions and bodily movements

I have said that there is a significant difference between the standard causal story outlined by Velleman (see 2.1.1, p. 12 above) and that suggested by Davidson (statements 2.2.1 and 2.2.2, p. 13 above). This is that Velleman cites reasons and *bodily movements* as the causal relata whereas Davidson cites reasons and *actions*. One possible way in which we might seek to avoid deviant causal chains is to give an account of what an action is, independently from its relation to reasons, and say that an example is only an example of acting for reasons if the reason causes an action, and not a mere bodily movement. This would help us avoid Davidson's climber counterexample (2.4.2 on p. 19, above). This does not give us an example of an *action* caused by a desire-belief pair, though it does give an example of a bodily movement caused by a desire-belief pair. However, if we are to capitalise on this distinction in order to protect the causal story from deviant causal chain counterexamples, we must demonstrate that bodily movements and actions are different. Davidson's comment that actions are 'mere movements of the body' (1971, p. 59) has sometimes been taken to suggest that he thought actions were bodily movements$_I$. In this chapter, I consider the plausibility of this stance, and will finally suggest that a characterisation of actions that is independent of their relationship to reasons must always be unsatisfactory. Given that I identified actions as things done for reasons in the Introduction, this should not be surprising. However, I will go on to argue that this conceptual link between actions and reasons does not rule out a causal theory of action.

## *3.1. Bodily movements$_T$ and bodily movements$_I$*

Hornsby argues that the statement "All actions are bodily movements" is ambiguous, and in one sense it is true (perhaps with a few exceptions) and in another definitely false (1980, Ch. 1 §1). The ambiguity arises from the linguistic fact that 'move' is one of a class of verbs that can be either

transitive or intransitive (i.e. it can occur with or without a grammatical object). I will follow Hornsby in using the subscripts '$_T$' and '$_I$' to distinguish transitive and intransitive uses of verbs respectively. A test of whether a verb belongs to the class with which Hornsby is concerned is whether it supports inferences from '*a* $V_T$ *b*' to '*b* $V_I$'. For example, 'melt' belongs to this class because we can infer from 'Jane melted$_T$ the chocolate' that 'the chocolate melted$_I$' (1980, p. 2). It is usually clear whether the transitive or intransitive form of the verb is being used from the context, but nominals formed from these verbs, such as 'the melting of the chocolate', can be ambiguous. It can be unclear whether it is the melting$_T$ of the chocolate (Jane's melting$_T$ the chocolate) or the melting$_I$ of the chocolate (the chocolate's melting$_I$) that is intended. 'Move' can disguise a similar ambiguity: between Jane's moving$_T$ her arm and Jane's arm's moving$_I$. Hornsby contends that it is implausible that all actions are bodily movements$_I$: if, when Jane has moved her arm, we ask what Jane did, it seems wrong to answer 'Jane's arm moved'. Hornsby does think that '[a]ctions can sometimes be identified with bodily movements$_T$' (p. 5).

## *3.2. Individuating actions*

However, as Hornsby points out, that bodily movements$_I$ and bodily movements$_T$ have different truth conditions for their occurrence and reports of each are answers to different questions does not entail that any particular bodily movement$_T$ is not the same event as a bodily movement$_I$. In order to show that they are different, Hornsby first turns to the discussion of the individuation of action. Two classic discussions are found in Anscombe (1957) and Davidson (1971). Anscombe's discussion focuses around the example as follows:

*3.2.1* A man, *x*, moves his hand in a certain way, *A*, which operates a pump, *B*, which in turn replenishes a water supply for a house, *C*, which in turn poisons a household, *D*.

Davidson's example is similar in structure and goes: An agent, $x$, moves his hand in a certain way, which flips a switch, which turns on the light, which alerts a burglar to his presence. In each case, four possible action descriptions suggest themselves. For 3.2.1 we might say:

A. $x$ moves his hand.

B. $x$ operates the pump.

C. $x$ replenishes the water supply.

D. $x$ poisons the inhabitants.

A similar list can be constructed for Davidson's example. Anscombe, Davidson and Hornsby contend that in each case there is just one action variously described as, e.g. a hand moving, a pump operating, a water supply replenishing and an inhabitant poisoning. This contention is disputed by Alvin Goldman (see, e.g. Goldman, 1970), who holds that A through D above refer to separate actions. Goldman points out that it is true that '$x$ poisoned the inhabitants by operating the pump', and argues that it is not true that '$x$ operated the pump by poisoning the inhabitants'. 'By', he argues is asymmetric and irreflexive and cannot hold between identicals. Any sentence of the form '$x$ $\phi$-d by $\psi$-ing' must, argues Goldman, report two distinct actions, a $\phi$-ing and a $\psi$-ing. Hornsby (1980, p. 7) replies that Goldman's formulation suppresses the fact that the sentence '$x$ $\phi$-d by $\psi$-ing' does not express a relationship between a $\phi$-ing and a $\psi$-ing. If this were so, then it seems that '$x$'s replenishing of the water supply was by $x$'s operating of the pump' should be true and '$x$'s replenishing of the water supply was by $x$'s poisoning of the inhabitants' should be false, whereas it seems that neither actually has any sense. She suggests that, rather than '$x$ $\phi$-d by $\psi$-ing' reporting a relation between an action of $\phi$-ing and an action of $\psi$-ing, it reports one action, perhaps a $\mu$-ing, of '$\phi$-ing by $\psi$-ing'. For example, 'he poisoned the inhabitants by operating the pump' should be thought of as reporting an action of '$x$'s poisoning-the-inhabitants-by-operating-the-pump', where the hyphenated phrase should be understood as a complex verb. She cites, as evidence

that 'he poisoned the inhabitants by operating the pump' does not mention any action of operating a pump, the fact that the verb 'operating' is not inflected for person or tense: 'they poisoned the inhabitants by operating the pump' and 'you will poison the inhabitants by operating the pump' are grammatically correct and not, for example, 'they poisoned the inhabitants by operated the pump'.

However, using a suspect 'by'-relation is not the only way to motivate the problem of individuating actions. George Wilson (1989) characterises two ways in which the problem can be raised. The first invokes the causal properties of the events in question: it seems that (i) '*x*'s operating the pump caused the poisoning of the inhabitants' is true, whereas (ii) '*x*'s poisoning of the inhabitants caused the operating of the pump' seems false. The second type of argument makes reference to the temporal properties of the events. Events have a more or less determinate duration, and it seems that the operating of the pump takes less time than the poisoning of the inhabitants, or even the replenishing of the water supply (presumably the water takes time to travel from the pump to the house).

However, Hornsby has already given us the resources to meet the first type of argument. The apparent problem raised trades on something akin to the ambiguity between transitive and intransitive uses of the verbs in question. I can operate the pump, and the pump can operate; I can poison the inhabitants, and the inhabitants can be poisoned. In the latter case, particularly, it is clear that the event of my poisoning the inhabitants and the event of inhabitants being poisoned need not take the same amount of time, nor even overlap in time: the inhabitants being poisoned would probably be taken to begin when the toxins entered their bodies, when they drank the water. Thus, whilst it is true that '*x*'s operating the pump caused the inhabitants to be poisoned', it seems false that '*x*'s operating the pump caused the *x*'s poisoning of the inhabitants', and that '*x*'s poisoning of the inhabitants caused *x*'s operating the pump'. Thus (i) and (ii) above can be seen not to simply reverse the causal relation between

two events of the same type. However, the question remains of how we should deal with statement (ii). It is clearly false that 'the inhabitants being poisoned caused the pump's operating', but Wilson and Hornsby both argue that it is in fact true to say that '*x*'s poisoning of the inhabitants caused the pump's operating', though this is admittedly not a sentence we would frequently use.

This implies (if we are not to admit cases of backwards causation) that *x*'s poisoning of the water supply occurred before the pump operated. This leads us to consideration of the temporal properties of actions, and so to the second type of criticism that might be levelled by the theorist committed to claiming that sentences A through D (on p. 33 above) refer to distinct actions. The argument is that the poisoning of the inhabitants takes longer than the operating of the pump, and thus they cannot be the same event. Again, Hornsby's distinction between transitive and intransitive uses of verbs is of service. The inhabitants being poisoned clearly does take longer than the operating of the pump, and probably does not even start until some time afterwards, when one of the inhabitants first drinks some of the contaminated water. But it is not clear that *x*'s poisoning of the water takes longer than *x*'s operating of the pump. It could be some time between the agent's operating the pump and the inhabitants' drinking the water, and in the meantime he could have done all manner of other things, such as read a book, had a nap, perhaps he even had a drink of the contaminated water before he pumped it to the house and dropped dead before the first inhabitant drank. Is it plausible to think that all this time he was poisoning the inhabitants? The case in which the agent is dead before the inhabitants drink is particularly stark. We do not usually suppose that someone can still be doing something after they have died. Rather, it seems quite natural to say the he poisoned the inhabitants *before* he died, and in fact, in operating the pump, he had *done* all he needed to do in order that the inhabitants be poisoned and the water supply replenished. Similarly, in moving his hand, *x* did all that was required (given the appropriate arrangement of

objects around him) to operate the pump. I therefore conclude, with Anscombe, Davidson and Hornsby, that A through D are in fact alternative descriptions of one and the same action, though it remains to be answered when exactly that action occurred and what relation the various descriptions have to each other and to statements using the intransitive form of the same verb that name different events.

### *3.3. Basic actions*

Above, it has been established that $\phi$-ing$_T$ and $\phi$-ing$_I$ are not necessarily the same event. Further, a phrase like 'the poisoning of the inhabitants' can be ambiguous between 'x's poisoning of the inhabitants' and 'the inhabitants being poisoned', i.e. the toxins entering their bodies and causing harm. Wilson calls this the 'act/process' ambiguity. It yet remains to be seen whether $\phi$-ing$_T$ and $\phi$-ing$_I$ are *never* the same event, particularly with reference to bodily movements, and I will return to this question shortly. Before that, I will briefly discuss Hornsby's account of the relationship between *x*'s poisoning and the inhabitants being poisoned, which I think is exactly right. I have argued that *x*'s poisoning is the same action as *x*'s hand moving, operating of the pump and replenishing of the water supply. Here, the hand's moving caused the pump's operating caused the water supply to be replenished which in turn caused the inhabitants to be poisoned. Hornsby's suggestion, that what makes *x*'s action, for example, a poisoning of the inhabitants is that *x*'s action set in motion a causal chain that resulted in the inhabitants being poisoned, thus seems a natural one to adopt. More generally, if an action of *x* is correctly describable as a $\phi$-ing$_T$, then *x* performed an action which caused a $\phi$-ing$_I$. Hornsby suggests that many action descriptions 'introduce' effect descriptions, for example, the action description '*x*'s operating the pump' introduces the effect description 'the pump's operating', and '*x*'s poisoning the inhabitants' introduces 'the inhabitants' being poisoned'. The effect descriptions thus introduced refer to events that are causally related, and some of these effects occur closer to the

time of the action than others: the operating$_I$ of the pump happened before the poisoning$_I$ of the inhabitants. Hornsby introduces the notion of basic$_c$ to describe this relationship:

*3.3.1* A description *d* of a particular action *a* is a more basic$_c$ description than another description *d'* if the effect that is introduced by (*d*,*a*) causes the effect that is introduced by (*d'*,*a*). (Hornsby, 1980; p. 71)[5]

Presumably, however, if we are to avoid a regress, this causal chain must be stopped somewhere, which encourages one to look for the *most* basic action, or simply *the* basic action. There must be a point where there was no action of the agent's which caused an event that was prior to the one under consideration

Danto's paper, 'Basic Actions' (Danto, 1965) has influenced much subsequent work on basic actions, but his account is unsatisfactory. Brand (1968) gives a concise and clear argument why. He characterises Danto's position as follows:

(D1) For every person *S* and every *a*, *a* is a basic action performed by *S* if and only if (i) *a* is an action performed by *S* and (ii) there is no other action *b* such that *S* performed *b* and *b* caused *a*.

(D2) For every person *S* and every *a*, *a* is a non-basic action performed by *S* if and only if (i) *a* is an action performed by *S* and (ii) it is not the case that *a* is a basic action performed by *S*.

However, as Brand points out, this entails that actions of ours cause other actions. While this may sometimes be the case (e.g. by leaving my keys in the house I might cause myself to pick the lock later, from Hornsby, 1980), it seems false to say that, for example, my operating the pump caused my poisoning the inhabitants. In fact, if our account of the individuation of actions is correct then this must be false, for if my $\phi$-ing

[5] It is not, though, the case that every description of an action introduces a separate event which is causally related to an event introduced by a different description. It could be that the effects introduced by two descriptions are identical, e.g. a description as a hand raising, and a description as a signalling (by raising one's arm). Thalberg introduces the notion of 'conventional generation' to describe what is going on in this sort of case.

causes my $\psi$-ing then my $\phi$-ing cannot be identical with my $\psi$-ing. Thus, like Brand and Hornsby, I reject Danto's notion of a basic action.

I will now look at Hornsby's proposal for the basic description of an action should be understood, and attempt to decide whether it is satisfactory. The discussion so far has traced the chain of events back to bodily movements$_I$, and I now intend to see whether it has to be traced back even further in order to identify the basic description of the action. Hornsby argues that, if a causal, rather than teleological, notion of basicness, then it must: she cites Wilkins (1668), who put forward a claim that Hornsby says is undisputed by linguists. This claim amounts to 'it is a necessary condition of the truth of '*a* $\phi$-s$_T$ *b*' that '*a* cause *b* to $\phi_I$'' (Hornsby, 1980; p. 13). Thus, Hornsby argues, bodily movements$_T$ cause bodily movements$_I$ and so can neither be nor be constituted by bodily movements$_I$. This, however, is too quick. If we consider '*x* moved$_T$ his body' and then use Wilkins' principle, we can see that '*x* caused his body to move$_I$' is suggested as a necessary condition of its truth. This is not the same as saying '*x*'s bodily movement$_T$ caused his body to move$_I$' is a necessary condition of '*x* moved$_T$ his body'. Therefore, Hornsby cannot use Wilkins' claim to demonstrate that bodily movements$_T$ are not bodily movements$_I$, or that they cause bodily movements$_I$. I am not sure that Wilkins' principle can be entirely correct, if it entails that statements like 'Jane caused her arm to move' are strictly true when 'Jane moved her arm' is true. Nonetheless, I expect any plausible theory to account for the plausibility of a statement like 'Jane caused her arm to move', and show how they are true, or nearly true. This will be one of the tests of how far I have succeeded in satisfying condition (4), the condition that specifies that a theory of action should give a plausible rôle for the agent in his actions. Hornsby thinks that '*x*'s bodily movement$_T$ caused his body to move$_I$' is part of a correct analysis of '*x* caused his body to move$_I$'. However, simple substitution of terms into Wilkins' principle does not demonstrate that this is correct.

Hornsby, however, supports her observations about Wilkins' principle with a more convincing argument for her claim that bodily movements$_T$ cause bodily movements$_I$. Thus far, we have looked at causal chains that stretch from the body outwards, but not at how we may affect changes within our bodies. Hornsby considers the apparently understandable claim that '*x* contracted his muscles by clenching his fist'. We are to suppose that a physiologist informed *x* that clenching his fist would accomplish the required muscle contraction and then asked *x* to contract his muscles in the appropriate way, and that *x* complied. Some theorists (Hornsby says that von Wright (1971) holds this view) have suggested that such examples give us instances of backwards causation. I agree with Hornsby, though, that this is an undesirable conclusion. Surely, whenever I clench my fist I contract my muscles, even if this is not my aim in clenching my fist, and I do not do so intentionally. Some philosophers might be reconciled to accepting that there is backwards causation in rare cases (though I do not include myself in their number), but it seems incredible to claim that there is backwards causation whenever anyone clenches their fist. The above formulation of our current problem again relies on the 'by' relation, which I have not analysed. In fact, Hornsby admits to suspecting that '*x* contracts his muscles by clenching his fist' is never true (Hornsby, 1980; p. 94), but proposes an alternative formulation of the problem. This trades on the claim that whenever I clench my fist I contract my muscles, and follows the earlier pattern of analysis of actions A through D (p. 33, above). Hornsby suggests that my clenching$_T$ my fist causes my muscles to contract$_I$, which does seem *prima facie* plausible. But if this is the case, and we are to avoid backwards causation, then it seems that *x*'s fist clenching$_T$ cannot be the same event as his fist's clenching$_I$: the muscle contraction$_I$ occurs before the fist clenching$_I$. Neither, Hornsby argues, can the fist's clenchng$_I$ be *part* of the fist clenching$_T$. She has two slightly different arguments to this conclusion. The first goes as follows (Hornsby, 1980 Ch. II §2.3):

(1) The fist clenching$_T$ is composed of the fist clenching$_I$.

(2) *x*'s contracting$_T$ his muscles is identical with his clenching$_T$ his fist, and therefore is composed of the same events.

(3) Therefore, *x*'s muscle contracting$_T$ is composed of the fist clenching$_I$.

The conclusion (3) is false, says Hornsby: 'the fist's clenching is not a part of his contracting$_T$ his muscles, at least if his contracting$_T$ his muscles is over by the time his muscles have contracted' (p. 23). I agree with Hornsby that an event cannot be composed of another event that lasts longer than it. Therefore, if we are to deny that the fist clenching$_T$ occurs before the fist clenching$_I$, we must deny that *x*'s contracting$_T$ his muscles is over by the time his muscles have contracted. Hornsby rejects this possibility on the grounds that, if a man with his hand amputated and replaced with an inflexible block connected to his muscles were to contract$_T$ his muscles in the same way, then there would be 'no event later than his muscles' contracting by which we could seek to fix the finish of his contracting$_T$ his muscles' (p. 23). But it seems that we could debate whether this man could in fact contract$_T$ his muscles in the same way: if he were to contract the muscles that we contract in clenching our fists it seems that he must go about it in quite a different manner and that therefore it could be a quite different action, not requiring the same temporal boundaries as our muscle contracting$_T$. He could not contract his muscles by clenching his fist. Thus, I suggest that there is no real reason to suppose that my contracting my muscles is over before my fist clenches, if my contracting my muscles is by clenching my fist. I hope my position will become clearer in the course of discussion of Hornsby's second argument.

This second argument is explicitly directed against Thalberg (1977). She says that, if fist clenching$_T$ results in muscle contracting$_I$, then clenching$_T$, *a*, causes contracting$_I$, *e*, which causes clenching$_I$, *f*, where *f* is part of *a*. This she says, sounds like a causal loop (p. 25), and I agree.

Thus, if we are to avoid Hornsby's conclusion that fist clenchings$_I$ are not part of fist clenchings$_T$ we must deny the antecedent and assert that fist clenching$_T$ does *not* cause muscle contracting$_I$. This, Hornsby objects, is counterintuitive, but I believe that further consideration might remove some intuitive distaste at this idea. The statement we must deny is that fist clenching$_T$ causes muscle contracting$_I$, the key move in making this plausible is to claim that the fist moving$_T$ is composed not only of the fist's moving$_I$, but also of the muscles' contracting$_I$, and of all the other physiological changes that occur when we embark on clenching our fist. Clearly, if the muscle contracting$_I$ is part of the fist clenching$_T$, then it cannot be caused by the latter. Rather, when we intentionally contract$_T$ our muscles, we make use of our knowledge of our physiology and clench$_T$ our fists in the confident knowledge that in the process of this action our muscles will contract$_I$. On this view, fist clenching$_T$ need not cause either muscle contracting$_I$ or fist clenching$_I$.

Admittedly, this looks more like a flat denial of Hornsby's view than a refutation of it, and in order to motivate my interpretation of it over hers I will have to bring further considerations to bear. In the process, I will have to explain why I advocate this different treatment of events inside the body than events outside my body, where I was completely convinced by Hornsby's treatment. One point in favour of this discrimination is that we recognise ourselves as having somehow direct control over our fist's clenching$_I$, but not our muscles' contracting$_I$. On Hornsby's view, the description of my action as a muscle contracting$_T$ *introduces*[6] the event of muscle contracting$_T$, the description of my action as a fist clenching$_T$ introduces the event of fist clenching$_I$, and the description of my action as, say, squeezing a rubber ball, introduces the event of, perhaps, the ball's becoming$_I$ deformed[7]. The muscle contracting$_I$ causes the fist clenching$_I$

[6] See §3.3, p.36 above, for an explanation of what 'introduces' means here.

[7] This illustrates the point made by Hornsby that it is not always easy to see how to describe the event introduced by a particular action description.

which causes the ball's becoming$_I$ deformed, and the action description that introduces each is successively less causally basic. This does not leave much room for the special relationship we might think we have with our fist clenching$_T$, or our fist clenching$_I$. Hornsby does remark on the fact that the things we do think that we have (direct) control over start at the surface of our body (1980, ch. 6 §3.2), but this observation does not play any part in her definition of causal basicness. She argues that the *most* causally basic action is one that has no more basic$_C$ description (see 3.3.1 on p. 37, above). That is, the description of the action such that the event it introduces is not caused by any event introduced by an alternative description of the action. She goes on to say that an action description that introduced no event at all would 'fit the bill' (1980, ch. 4 §3.2). If it were not an action description that introduced no event at all, the most basic description of the action would be, perhaps, as an altering of my brain's state, and this seems implausible. But Hornsby argues that there is a more basic description than this, one that introduces no event at all, namely a description of the action as a trying. However, it seems to me that the idea that the basic description of an action should introduce no event is wrong. That an action has taken place *depends* on some event's, such as a body's moving, having taken place, and that the most basic description of an action does not entail that any action has taken place seems deeply problematic. If Hornsby is right, then a satisfactory and useful causal notion of basicness is unavailable, which would be highly problematic for a causal theory of action. This worry, is at the root of my resistance to Hornsby's arguments so far. I will therefore attempt to identify an alternative account of the most causally basic description of an action, and as I have already hinted, I think that a description in terms of something we directly do is the most plausible candidate. Hornsby's account has further merits, however, and an investigation into how her discussion proceeds might reveal that these merits outweigh the problems identified. I will conduct such an investigation in the next chapter, and though I will conclude that I cannot

accept Hornsby's theory, consideration of it will stimulate ideas that will dictate my answer to the question of how bodily movements$_T$ are related to bodily movements$_I$, and how we should identify basic actions.

# Chapter 4. Trying

The primary purpose for my continued consideration of Hornsby's view is that she identifies actions with tryings. If this can be fitted into a general causal theory, then it seems that the way is clear to satisfying condition (4), by giving a plausible account of the rôle of an agent in his actions, and so of intentional actions. This is because, whenever I try to turn on the light, I know that I am doing so, and would probably also know my reasons for so doing (I want to be able to see, and believe that turning on the light will allow this, for example). This sort of awareness of what I am doing, and why I am doing it, is just what I identified as lacking in cases where I act for reasons, but not intentionally, and where, therefore, the agent was only minimally involved in his actions. However, I have already pointed out that a description of an action as a trying may be problematic when considered as the most basic description of an action because such a description does not entail that an action has taken place.

## *4.1. Actions and tryings*

Both O'Shaughnessy and Hornsby argue that, whenever there is an action, there is a trying. It has, however, been suggested this is false. We generally only use locutions like '$x$ tried to $\phi$' when there is at least the possibility of failure. It might seem, therefore, that 'tried to' is simply a marker to indicate our doubt that the agent will be successful in his attempt to $\phi$. This opinion has been convincingly argued against by O'Shaughnessy (O'Shaughnessy, 1973). He uses the device of a sceptical but rational observer to illustrate his point. He says that for any situation it is possible that such an observer might doubt the outcome of an agent's attempt, perhaps as a result of misinformation, and that the observer can know the agent will try, but doubt he will succeed. And if this observer *knows* it, it must be true. For example, if the observer knows an agent, $x$, has strong motives to start his car, but wrongly but rationally believes that this is not possible (perhaps he has been misinformed by a usually

reliable source), then the agent cannot be said to know that *x* will start his car, but can be said to know that *x* will *try* to start his car. It must always, argues O'Shaughnessy, be possible to envisage a situation in which an observer might rationally expect failure, so the link between trying and the possibility of failure is not broken. With Hornsby, I doubt that such a case can necessarily be constructed for *every* action an agent undertakes, but O'Shaughnessy demonstrates that the scope of cases in which agents try is much wider than we might initially suspect. However, I doubt that whenever there is an action there is a trying for a different reason. I do believe that there is a trying whenever something is done intentionally, but not whenever something is done for reasons. Consider the example given in the Introduction, where I lit a cigarette without realising what I was doing. I argued that my lighting a cigarette was an action, but not an intentional one. I also think that it is false to say that I *tried* to light the cigarette. A similar observation applies to Velleman's example, where he alienates his friend as a result of a reason of which he is unaware. Again, I do not think it is accurate to say he was trying to alienate his friend. It is true that we do sometimes speak of someone trying to do something without realising, for example, if the agent seems to have a 'death wish' we might say he is trying to kill himself, even though he does not realise it. I will argue later that we might say the agent is involved in his action, though not fully, if it is caused partly by a central goal of the agent, something that guides much of his behaviour, as a 'death wish' would do. On the other hand, we would also probably say that the agent intended to kill himself, but did not realise it, which counts against my definition of intentional action. It is to be hoped that a good theory of action would give us the resources to explain this sort of problem case, but, in the meantime, it does seem that the possibility of describing an action as a trying is an indicator of the involvement of the agent in his actions.

Before proceeding to discuss O'Shaughnessy's and Hornsby's views in more detail, I will briefly consider whether an analysis of actions as

partly, or wholly, tryings will provide the answer, not only to the question of how an agent is involved in his actions, but also immunise the account to deviant causal chain counterexamples. It might be thought it does not, and if tryings are actions and the *causes* of the events that entail an action has taken place, then I think that it certainly does not. Suppose Sally has been in hospital following an accident, and is very bruised. She tries to get up, but her moving her leg causes a pain which in turn causes her to shoot out of bed. Here, though Sally tried to get up, and did indeed end up out of bed, she did not get out of bed intentionally.

### *4.2. Actions as tryings plus bodily movements*$_I$

O'Shaughnessy thinks of an action as a composite event comprising both, for example, a trying and an arm's rising, whereas Hornsby thinks of an action as a trying alone. For Hornsby, there can be two type-identical events *a* and *b*, where *a* is an action and *b* not. What makes *a* an action what it causes. She suggests that we see this as a sort of mirror image of a causal account of perception, where a particular sense-datum counts as a perception only if it has the appropriate causal antecedents. Hornsby's account is partly a response to the problems she finds with O'Shaughnessy, and which I will shortly agree are insurmountable. However, because O'Shaughnessy thinks that actions include bodily movements$_I$, it seems his account could give an easy solution to the Sally example in the previous section, that is not available to Hornsby.

O'Shaughnessy says 'all physical action involves a willing or bringing about of act-neutral bodily events' (p. 369). Thus starting a car involves my bringing about the car starting, my raising my arm involves my bringing about the rising of my arm, and my walking from A to B involves my bringing about a series of scissor-like movements of my legs that propel me along from A to B (all examples from O'Shaughnessy, 1973). O'Shaughnessy identifies trying, bringing about and willing. However, I will not be discussing willing in this thesis. The question now must be what it is that I do that brings about the act-neutral event. For

O'Shaughnessy, that is trying to start the car, raise my arm or walk from A to B. So, for every physical action there is a trying that brings about the act-neutral event which has the same physical attributes as the action.

O'Shaughnessy asks us to:

*4.2.1*

> consider the case of a man who believes but is not quite certain that his arm is paralyzed; suppose him asked at a signal to try and raise his arm. At the given signal he tries, and to his surprise the arm moves; but a moment later he tries again and thinks he has succeeded, only to discover on looking down that he has failed. (O'Shaughnessy, 1973; p. 371)

Let us call the case where he tries and succeeds $T_S$ and the case where he tries but fails $T_F$. He argues that the same kind of trying-event takes place in both cases and that in both cases trying is an event: it is located in space and time, and has causes and effects. Further, he argues that trying is an action in both cases. Trying, he argues, shares the following attributes that are definitive of actions: it (a) is an event, (b) comes as no surprise to the agent, (c) happens because he intends it to, (d) originates in his reasoned desires and (e) sometimes trying to do something might be doing a physical action that will, world permitting, cause the desired outcome. He goes on to argue that, not only is there a trying that is an action in each instance of the intermittently paralysed patient's trying to raise his arm, but that the trying is the same kind of action in both instances. His conclusion is based on the consideration that 'trying to raise his arm' is what he calls the '*ultimate psychological description*' (p. 373), and applies in both cases. From the considerations that (1) the same kind of trying-event occurs in *TS* and *TF* and (2) the arm rising occurs in *TS* but not in *TF*, O'Shaughnessy concludes that (3) the 'trying and arm rising must be *two* events and two *distinct* events' (p. 373). This reasoning appears valid, but we need to know the connection between the two events. O'Shaughnessy argues that the connection is explanatory, insofar as 'trying both explains and is a sufficient condition of arm rising, in normal physicopsychological circumstances. Therefore the event, trying to raise the arm, normally *causes* the event, arm rising' (p. 374). His

conclusion is that 'the logically necessary requirements of an intentional act of arm rising are: a causally linked pair of events, trying and arm rising' (p. 375).

At this point, it seems natural to wonder why O'Shaughnessy thinks that the arm's rising is part of the action of arm raising. He has argued that the trying implicated in the action is the same in cases *TS* and *TF*, and, what is more, it is an action in its own right in each case. This implies that, each time we act we perform two actions: one that is the trying alone, and another act that is composed of the trying-act and a bodily movement. O'Shaughnessy, however, both wishes to avoid this consequence (as he believes it has dualist implications), and thinks that he can. The passage in which he attempts to justify his position is difficult enough to deserve quotation.

> The inner event, trying, though merely internal and therefore lacking absolute guarantee of its object in the outer world, has no other existence that as a putative cause of a bodily event. Were the act of trying *independently specifiable*, an act like the imaging of a bodily event, then whereas the act would putatively relate to items in the outer world and also putatively be a cause of arm rising, it would retain its identity as a distinctive act of a certain kind. We would then merely *redescribe* the act as "trying to raise the arm," and its causal efficacy would be a simple contingent property akin to that of the thought of a steak to cause watering of the mouth. (O'Shaughnessy, 1973; p. 378).

O'Shaughnessy attempts to bind the trying and the physical event together in such a way that it is plausible to say they constitute a composite event, namely a physical action. He seems right to argue that there would be little justification for including the physical event in the action if the trying were specifiable independently, was an act in its own right. However, we generally think that causal connection *has* to be contingent, and that the causal relata involved must therefore be independently specifiable. Davidson argues (e.g. 1963) that this need not be the case at some particular level of description, but that it is enough that there is *some* description of the events such that they are

independently specifiable. His idea is that there are events which can be described in any number of different ways, and which might be in singular causal relationships independently of any particular description. Any particular causal explanation does not require the relata to be logically distinct. Unlike Davidson, I do not think that events are completely description independent, and I do believe that, if an explanation is to be causal, then the events mentioned must be logically independent as described in that explanation[8]. Unfortunately, a full discussion of this point is outside the scope of this thesis, but my views are similar to those expressed by Hornsby. I therefore conclude that O'Shaughnessy's position is untenable, and that an alternative view must be sought.

### *4.3. Actions as tryings*

I have dismissed the idea that physical actions are tryings plus bodily movements, but have not yet produced an objection to O'Shaughnessy's claim that trying is a species of action identifiable in all cases of physical action. Specifically, I have not yet considered how we might escape the consequences of his argument based on the example of the intermittently paralysed man, who tries to raise his arm twice, once succeeding and once failing, but, O'Shaughnessy argues, doing the same thing each time. Without a reply to this argument, which purports to tell us what tryings are, but unable to accept that this allows us to specify that physical actions include bodily movements$_I$ as well as tryings, we are left with the option that actions are tryings, and *only* tryings. This is Hornsby's contention. She says, 'Every action is an event of *trying* or attempting to act, and every attempt that is an action precedes and causes a contraction$_I$ of muscles and a movement$_I$ of the body' (1980, p. 33). Hornsby argues that all actions are tryings, and, though not all tryings

[8] Though, as long as we are clear about the strictly true causal explanatory statements, which describe events in a fairly similar way to an explanation that describes events in such a way that there is some logical entanglement, we might not say that this latter explanation is strictly false. c.f. §5.5 p. 76, below.

are actions, those that are cause muscle contracting$_I$ and movements$_I$ of the body. This would explain why we feel events like bodily movements$_I$ are so central to actions: they do occur whenever there is an action. Thus Hornsby's suggested analysis of an action is a trying that causes physical movement.

However, I said that O'Shaughnessy's motivation for including physical events in actions was that of avoiding dualism, and so we must see how Hornsby would overcome this difficulty. There is a way in which Hornsby's actions are intimately connected with the physical world, and that is, that if a trying is an action, then it has physical effects. I am not sure, however, that this is enough. It is widely accepted that desires can have physical effects, yet this does not mean that desires are physical, and that we only *call* tryings actions if they do have physical effects does not seem to capture sufficiently idea that physical actions are essentially physical. I will turn my attention to this shortly. First, however, I will argue that it is not enough merely to show that actions are not mental. Hornsby does admit that actions are 'inside the body', and this raises problems in itself. O'Shaughnessy (1973) did consider the possibility of identifying actions with tryings, rather than tryings plus bodily movements$_I$, but concluded that this identification was undesirable on the grounds that this 'implies that the act of raising the arm causes arm rising, and so it follows that all the evident physical effects of the act must be mediated by arm rising' (p. 385). If my arm rising knocks over a vase, for example, it would be the case that there was a stretch in the causal chain between my act and the vase falling over that was opaque, even though I saw my elbow hit the vase. As a result, actions cannot be thought of as directly causing physical events outside my body, which seems implausible.

I think that O'Shaughnessy's argument gives us sufficient reason to worry that, if tryings are inside the body, then actions cannot be tryings. However, the charge that, on Hornsby's view, actions are not merely inside the body, but actually mental, is still more serious. Hornsby does

not devote much space to the charge that her position is dualist in *Actions*. She simply says:

> Any dualist who wants to introduce a purely mental item into my account of action would need to look back beyond the trying (which is the action, and is something we know is physical). (Hornsby, 1980; p. 59)

Her argument seems to be that, the action is physical, the trying is the action, therefore the trying is physical. Though she no doubt has a more subtle point in mind, what she actually says here fails to address any concerns we might have that Hornsby's account leaves neither actions nor tryings as physical. Hornsby needs more of an argument to show that her analysis has not pushed action into the realm of the mental. Since *Actions*, she has addressed the concerns philosophers have expressed on this score. In a forthcoming paper, she argues that the mistake of seeing her view of dualist rests on faulty naturalistic assumptions. She says that the problem 'arises from supposing that the bodily movements that there are when there are action might be located in a world bereft of beings who do things for reasons'. To the contrary, she says, 'There need be no unitary category to which movements that there are when there are actions and what are called "mere movements of bodies" all belong' (Hornsby, forthcoming). Given, however, her view of actions as tryings that cause physical events, I am not sure that this response is available to her. Hornsby argues that the statement '$x$ contracted$_T$ his muscles by clenching$_T$ his fist' shows that there is some event $a$ that causes both the contracting$_I$ and the clenching$_I$. But suppose, rather than the physiologist telling $x$ that, if he clenches his fist, then the required muscle contraction will occur, then asks us to contract our muscles in that way, he tells him that, if he clenches his fist, then a required brain event will occur, and asks $x$ to bring about this brain event, which $x$, by clenching his fist, does. It seems that we could repeat this thought experiment for every event in the causal chain leading from the start of the action to the fist's clenching$_I$, and that the trying must be the cause of all of them. Actions, therefore, are well inside the body. However, the charge of dualism has

not yet been sustained. The way to make this final step , it seems, is to argue that, on Hornsby's account, trying cannot be a physical event, or composed of any physical events, at all. If her account of contracting one's muscles by clenching one's fist is correct, then the possibility of such a locution implies that there is an event, the trying to clench one's fist, that is the cause of both the muscles' contracting and the fist's clenching. But, in that case, any physical event that was suggested as identical with the trying or part of the trying could be brought about by the agent by clenching his fist, and is therefore caused by the trying, and is not identical with it or part of it. Thus, it seems to be Hornsby's conception of trying as the cause of all the physical events whose occurrence makes it an action that chases it back into the realm of the mental. Her contention that the physical events that entail an action has taken place are themselves not part of the action does not seem to allow us to justify any claim that they are not in the same category as physical events not brought about by any exercise of agency. I think that the solution is to argue, not that there is not unitary category that bodily movements that are or are not the effects of tryings/actions belong, but that there is no unitary category to which tryings that are and are not actions belong.

## *4.4. Tryings as actions*

I have suggested that both O'Shaughnessy's and Hornsby's accounts are problematic. On the other hand, I am convinced by their arguments that some tryings are actions, for example, that my trying to start the car is my turning the key in the ignition. In this section, I want to explain what I think tryings are, in such a way as to avoid the problematic regress into the body and possibly the mental threatened by Hornsby's account, but which does not entail that physical actions are composed of tryings and physical events. The key move is to deny O'Shaughnessy's and Hornsby's interpretation of the example of the intermittently paralysed man. It is worth restating here.

*4.2.1* consider the case of a man who believes but is not quite certain that his arm is paralyzed; suppose him asked at a signal to try and raise his arm. At the given signal he tries, and to his surprise the arm moves; but a moment later he tries again and thinks he has succeeded, only to discover on looking down that he has failed. (O'Shaughnessy, 1973; p. 371)

As before, I will call the instance in which he succeeds *TS*, and that in which he fails *TF*. Both O'Shaughnessy and Hornsby argue that there is a trying event common to both *TS* and *TF*. I, on the other hand, suggest that tryings that are actions are in a different category from those that are not. In O'Shaughnessy's example, the two events are phenomenologically indistinguishable by the agent, but disjunctive theories of perception give us a precedent for denying that phenomenological indistinguishability is a guarantor of two events being of the same kind. This approach allows an account that claims actions are tryings to shake off accusations of Cartesianism. Tryings that are actions necessarily are composed of physical movements. The question that must now arise is what relationship I envisage between tryings that are actions and bodily movements$_I$. In a sense, I think that actions, and so tryings that are actions, are composed of bodily movements$_I$. I also do not think they are composed of anything else. As I have said, I do not think there is a distinguishable event of trying common to instances where the trying is not an action and where it is, and I agree with Hornsby that, other than bodily movements$_I$, there is no plausible alternative to trying as a component of action. Rather, I think that tryings that are actions are also, in a sense, composed only of bodily movements$_I$.[9] But I do not think that actions are mere bodily movements$_I$.

This will need some explanation, and in order to provide it, I must marshal the arguments from the previous chapter and this one. In Chapter 3, I argued against Hornsby's conclusion from the that an agent's

[9] This is not to say that there cannot be a number of phenomenological features that commonly occur when we try to do something, such as a feeling of effort or strain. These features might be present whether we succeed in acting or not.

contracting$_T$ his muscles could be identical with his clenching$_T$ his fist our muscles, but that the contracting$_I$ of the muscles caused the clenching$_I$ of the fist that the action (the contracting$_T$ and the clenching$_T$) must have happened before both the contracting$_I$ and the clenching$_T$, and have caused both. I hinted that the action should be considered to last as long as a basic action, and that the notion of actions directly performed would give us the appropriate conception of a basic action.

There are a number of things that we consider ourselves as able to directly do. Many of these, though not all, are movings of our bodies, for example, clenching our fist. We can also, for instance, breathe more deeply and slowly. We do not usually control our breathing, and would probably not call our ordinary breathing an action, but it is a fact about us that we can often take control of this process. When we intentionally breathe more deeply there will probably be effects on the surface of the body, such as our chest rising and falling to a more pronounced extent, but, unlike the case of our contracting our muscles by clenching our fist, it is wrong to say that we breathe more deeply by moving our chests in a particular way. It seems to be a fact that some people are able to control their bodies to different extents. Some people, for example, can wiggle their ears, whereas I would not even know how to go about trying to do this. Some can fold their tongues, whilst others cannot. It also seems to be the case that we can learn to directly control our bodies in ways that we previously could not, such as those people who can allegedly directly slow their heart rate.[10] Certainly, early infancy appears to be a time during which we learn to extend our control of our bodies. We can also perform indirectly actions that we could do directly (for example, I could raise my right arm directly, or by raising my left while it is under my right). I take the things that we can directly do to be given independently of a theory of

[10] It seems plausible that many people who can alter their heart rates do so indirectly, but *via* a mental action, such as thinking of a calm place, rather than *via* another physical action. That some mental actions seem to have physical effects is very interesting, but falls outside the scope of this thesis. I will only consider physical actions and their effects.

action. The explanation of what things they are is possibly in the realm of evolutionary biology, and also, no doubt, the tings we can do directly are limited to things that make changes that we can detect. Hornsby says:

> Presumably for an organism to have control, it must be in a position to make changes that are contingent on changes it can detect. (Perception is the handmaid of action.) (Hornsby, 1980; ch. 6 §3.2)

My suggestion then, is that a basic description of an action is one which describes it as something directly done by the agent, which is as a direct exercise of control by the agent of his body[11] (though not necessarily a movement of the surface of his body). This conception of a basic action does entail that an action has taken place: it introduces an effect (an arm's rising, perhaps). I will call the effect that a basic description of an action introduces a 'basic movement' (which is a movement$_I$, and might not always be a movement on the outside of an agent's body, but a change in some part of the agent's body). I do not think a basic action of this sort causes the basic movement, the description of the action as basic (in my sense) is not logically independent of the basic movement (as the description entails that the movement has occurred), and the two descriptions (of basic action and basic movement) are not, therefore, suitable for inclusion in a causal explanatory statement of the form 'the basic action causes the basic movement'.

This conception of a basic description of an action can help clear up a mystery that results from Hornsby's account, that is, how we decide when a trying is or is not an action. Hornsby considers the case of Landry's patient, which was discussed by William James (1890). This patient had no proprioception of his arm, but could raise it when asked to do so. If,

[11] It may be that moving our bodies is not all we directly do. Baier (1971) has suggested that we also directly speak sentences, and this seems quite plausible. Our ability to make noise is certainly something that comes quite naturally to us. It is significant, I think, that all we need to do in order to speak a sentence is move a bit of our body. Annas (1978) has argued that it is also possible to directly type the letter 'p'. It does seem the case that we attain such mastery of machinery, like computers and cars, that they seem almost like extensions of our body. I think that it is possible to directly control things that are not parts of our bodies, though I believe that our control over our bodies is a prerequisite of the extension of this control. However, the question of where the limit of things we can directly do lies is outside the scope of this work, and its answer does not materially affect my thesis.

however, his arm was restrained without his knowledge, and he was asked to raise it, he would try to do so, and be quite surprised to discover that he had failed. I propose to agree with Hornsby that there is no action when the patient fails to raise his arm, even though there might be muscle contractings$_I$, and an upward force exerted on the restraint. However, if Hornsby's account is right, this is a case of a trying that has physical effects, and she owes us an explanation of why it is not an action. She says, talking about Landry's patient:

> But it must be remembered that the patient intended to do no more than move his arm. He did not set out to tense it, or to contract his muscles. Of course if the muscles do contract, that will be excellent evidence that he has tried to move his arm. But unless his arm also moves$_I$, there is no evidence that he moved$_T$ it. Even if all actions are tryings, not all tryings are actions. (Hornsby, 1980; p. 42)

Hornsby is responding here to the objection that the agent is the ultimate authority on whether or not he has acted, and so it is possible that she would respond differently to my objetion. However, it seems that she dismisses the idea that tensings$_I$ or contractings$_I$ are important in deciding whether the agent acted or not too swiftly, particularly when considering one of her previous arguments. When arguing that von Wright was wrong in thinking that there were some unusual cases, such as my contracting$_T$ my muscles by clenching$_T$ my fist, where there is backwards causation, she suggested that would entail that there is backwards causation whenever I clench my fist, as, whether I am conscious of it or not, or intend it or not, whenever I clench$_T$ my fist, I contract$_T$ my muscles. Extrapolating to the current case, it would seem that, while, in virtue of the patient's arm not moving, he did not move$_T$ his arm, he did contract$_T$ his muscles. We cannot say, as Hornsby seems to imply, that there is an action only if the agent accomplished what he intended. A problem in the causal chain of events inside the body entails that there is no action, in the way that a problem outside the body does not. For example, if I tried to turn on the light, but failed because the bulb was dead, we would still say there was an action (which could be

described as my trying to turn on the light, my moving my hand, or my flipping the switch). How is this different from Landry's patient, who does (probably) contract his muscles, but still not succeed in performing an action? I suggest that we need the account of basic actions that I have suggested in order to clear up this problem. An action is said to have occurred if an agent has succeeded in directly exercising control of his body, no matter what happens afterwards, but not otherwise.[12] However, if an agent succeeds in exerting a force on restraints of which he was unaware, then I would probably suggest that he has acted: the upwards force is evidence that he has succeeded in controlling his body.

## *4.5. Conclusion*

In this chapter, I have examined closely the idea that actions are, partly or wholly, tryings. The move was appealing because it seems that trying essentially involves an agent: it is an agent who tries to do something, and if he tries to do something, then he knows he is doing so, and usually knows why. Inclusion of trying in the theory therefore seemed to be a possible way to meet condition (4), that of giving a plausible rôle for the agent, and so to provide an account of intentional action. However, I have argued that conceiving of trying as the cause of the bodily movements$_I$ that a description of the action as, for example, an arm raising, introduces (in Hornsby's sense) does not resolve our problem. It is still possible to construct deviant causal chain counterexamples, and we run a serious risk of chasing both action and trying back into the realm of the mental (on a par with desires), or at least inside our bodies.

[12] I am not sure that Landry's patient does not act in the example given. As I have said, our proprioception is very important in determining what we can directly do, and Landry's patient lacks such proprioception of his arm. Therefore, what he can directly do, if anything, with his arm is undoubtedly very different from what those of us who do have proprioception can directly do. I do not think that the example of Landry's patient is particularly enlightening when considering action for the majority of us. People who lose proprioception, especially in their whole body, are often unable to move for some time, if ever. They report themselves as having to learn how to move again, and their reports of how they succeed in doing so reveals that their actions are undertaken in a quite different way from people who do have proprioception. Though we would expect a general theory of action to be able to explain all action, I do not think that considering cases as radically different from the more usual sorts of action undertaken by the majority of people at this early stage is particularly productive.

In response, I have argued that tryings are not mental, and that they are actions. The way to get this plausible result, I have suggested, is to deny that tryings that are actions fall into the same category as tryings that are not actions. I have further claimed that tryings that are actions are ones that are describable as basic actions, where a basic action is one we directly do. I have also introduced the ides of a basic movement, which is the effect introduced by the description as a basic action. However, I still owe the reader an explanation of how basic actions are related to basic movements. I have denied that the former is the cause of the latter. My intuitions suggest that an action, such as an arm raising, lasts until my arm stops rising. By denying that there is an identifiable 'trying' part, common to actions and non-actions, I have blocked the possibility of my saying that actions are bodily movements plus something else. Must I then, make the implausible claim that basic actions *are* basic movements? This question will be answered in the next chapter.

Before I am in a position to finalise the form my theory will take, I must consider whether descriptions of actions as tryings are the sorts of description that should appear in the theory. I conclude they are not. I have indicated that I do not think that to describe an action as a trying is to give the most basic description of it, but have not said what I do think we do when we so describe an action. I think that, when we describe an action as a trying, we describe it in terms of what the agent intended to achieve by so acting. Hornsby's concern was with the causal consequents of action, and therefore this feature of descriptions as tryings did not concern her. However, I stated my aim as that of finding a causal basis for the relationship between reasons and actions. A description of an action that involves essential reference to the agent's reasons is unsuitable for this purpose, because events related by a causal-explanatory statement must be logically distinct as so described. This is not the case for events described as reasons (or coming to have reasons) and events described as tryings. This raises an important point. Describing an action as a trying makes it explicit that the agent was

doing something intentionally, but I argued in the Introduction that actions are things done for reasons. I have not succeeded in finding an alternative characterisation of actions that does not make essential reference to the fact that the agent had reasons for so acting. I have argued that an action is more or less contemporaneous with a bodily movement (and perhaps with events inside the body that cause that movement), but have balked at claiming that these events in and on the body are identical with the action. I think it might be possible to argue that basic actions are basic movements, and that, if they are any of the events cited in causal relation in my theory, then they are bodily movements. However, I do not think that actions need be events actually mentioned in a causal theory of action, as long as the theory makes clear what happens when actions do. I will explain this further as I formulate my theory in the next chapter.

# Chapter 5. A general causal theory of action

At the end of Chapter 2, I argued that a general causal theory must:

(1) Be causal.

(2) Give an account of what an action is.

(3) Give some sort of directness constraint.

(4) Give a plausible rôle for the agent.

(5) Admit of degrees.

The purpose of the last two chapters has been to investigate how a theory might attempt to satisfy condition (2) in such a way to provide clues as to how conditions (3) and (4) might be met. I have largely ignored condition (5) thus far, and whilst causation has featured frequently in the discussion, the subject has mainly been events caused *by* actions, not the causation *of* actions. The question that motivated much of Chapter 2, regarding the relationship between reasons and actions, therefore remains to be answered.

## *5.1. Basic actions and basic movements*

I agree with Hornsby that an alternative characterisation of actions as tryings is the only plausible candidate for a redescription in terms other than as mere bodily movements. However, I argued in the last chapter that such a re-characterisation is not, ultimately, satisfactory for the purposes of constructing a general theory of action. To describe an action as a trying, I have maintained, is not to give a more basic description of an action, but to describe it in such a way as to link it to the agent's reasons or intentions, thus frustrating attempts to give a causal account of the relation between reasons and actions. Though describing an action as an attempt to achieve some aim is one way of explaining how it relates to the agent's reasons, the motivation behind this thesis has been the intuition between reasons and actions is causal, not only conceptual. I

have argued that the basic description of an action is as a direct exercise of control by the agent of his body. I take it that the actions we can directly perform are fixed independently of the theory of action. They are simply facts about us, though they may differ from agent to agent, and for one agent over time. Further, that we can bring something about directly does not entail that, whenever we bring it about, we do so directly. I might raise my right arm directly, or by moving it with my left arm. My rejection of tryings as appropriate descriptions, might seem to leave me with only the option of identifying basic actions with bodily movements. I do think that the action lasts about as long as the basic movement (the effect introduced by the basic description of the action). I am not sure exactly how long either lasts. I find it reasonably plausible that 'my arm rising' includes various causal antecedents (such as muscle contractings) of the outwardly visible effects of my arm rising. However, we will see that it does not much matter to the substance of the theory how long either takes, or whether they take the same amount of time.

I am trying to answer Velleman's question 'What happens when someone acts?' which I have now narrowed down to the question, 'What happens when someone moves$_T$ his body?'[13] My aim is to sketch the causal sequence that underpins our ability to describe an event as someone moving$_T$ their body, or acting. The answer that I have in mind is along the lines of 'When someone acts, a primary reason of theirs directly causes a basic movement.' Hornsby argues that the answer to the question 'What event on the sequence is the action to be identified with?'[14] should not be taken to determine '[a] theorist's overall view of the nature of *action*' (Hornsby, forthcoming). I agree wholeheartedly with this point, and suggest that, if actions are not identifiable with any event in the causal sequence I am outlining, then at least the causal sequence tells us what happens when an action takes place, that is, 'what happens when

[13] Or, rather, controls his body directly.

[14] From Brand (1984, ch. 1), quoted in Hornsby (forthcoming)

someone acts'. I argued that actions are things done for reasons. I now claim that actions occur whenever reasons cause movements in the right way, that when we do something for a reason, our reasons cause the effect introduced by the action description. More specifically, I will argue that an action takes place when a primary reason of an agent directly causes a basic movement, that is, an effect that is introduced by the description of the action as direct control over his body by the agent. I will talk more about the sorts of things primary reasons must be, and what it means for them to *directly* cause basic movements in the next section.

It is worth noticing, however, that I have abandoned any attempt to give an account of what an action is (other than to say that it is something done for a reason), and thereby satisfy condition (2), before formulating the theory. The theory I am proposing will not explain action by fitting it into a causal explanatory structure, but by showing how it is underpinned by such a causal-explanatory structure.

## *5.2. Directness*

A basic movement, I suggest, is what is directly caused by the agent's primary reason. I should now turn my attention to condition (3), giving a directness constraint. Davidson's definition of a primary reason for *A*-ing is, roughly, a desire for some end *E*, and a belief that performing some action *A* will result in *E*. To formulate the directness constraint, I will have to use Hornsby's idea of an event description *introduced* by an action description, discussed in §3.3 on page 36 above. Let us call a basic description of an action *BA*, and the basic movement that it introduced $BA_M$. I wish to specify that a primary reason is a reason for a basic action: a desire for *E*, plus a belief that *BA* will result in *E*. Whenever there is an action, I suggest, there is some true statement of the form 'a primary reason for *BA* directly caused $BA_M$'.

By 'directly', I mean that there are no intervening links in the causal chain between the primary reason and the basic movement. This may

initially seem to be implausible, after all, I have already discussed at length the idea that, whenever I clench my fist, my muscles contract which in turn causes my fist to clench. Even if the muscles' contracting is included in the basic movement, it might be that there are brain events that are caused by the primary reason, and which cause the first brain event that could plausibly be considered to be part of the action. However, in denying that there are intervening links in the causal chain, I am not denying that there are physiological, or neurophysiological and physiological, reductions of the causal process. My contention is that there are no intervening causal links when the events are described as psychological states or events, actions or basic movements. Descriptions of events as brain events or muscle contractings (unless they are ones I can directly perform) have no place in the sort of explanation I envisage. This rules out the letting go of the rope by Davidson's climber: in that case shock or horror intervened between the primary reason and the movement. The example described by Chisholm, where the nephew accidentally runs over a pedestrian who turns out to be the rich uncle the nephew had reason for killing, is a little harder to deal with. I have suggested that the running over the pedestrian could be thought of as an action as a result of the nephews intentional driving. But the nephew's driving was not caused by a reason for killing his uncle, and his bad driving was caused only indirectly, *via* shock, by a reason for killing his uncle. Thus, the reason for killing the uncle has no direct bearing on the question of whether or not running down the pedestrian was an action of the nephew's, though we might take into account his shock when considering how far he was in control of his driving when he ran down the pedestrian, and therefore how far it can be considered to be an action.

I have said that there are things, such as raising our left arm, that we can do directly, or by doing something else, perhaps raising my right arm when it is under my left. This might appear to raise difficulties for my directness constraint. I might have a primary reason for raising my left arm, and a primary reason for raising my right. I raise my right arm,

which causes my left to rise also. Here, my directness constraint rules out the raising of my left arm as a basic action, as it is caused by something else, the rising of my right, that can be described as a basic movement. Imagine, however, that I could clench my fist either directly, or by contracting my muscles. I think it probable that there are genuine examples along these lines, though I cannot think of any that are true of me. If it is not possible, then the problem does not arise, as 'a muscle contracting', for example, would not then be a description of a basic movement. But, though I cannot directly contract my muscles in such a way as to make my fist clench, if I could, my directness constraint would seem to suggest that, whenever I clench my fist, my basic action is always my contracting my muscles, even when I would say that I clenched my fist directly. If, as in the example discussed earlier, a doctor asks me to contract my muscles in the required way, then I have primary reason for contracting my muscles ($PR_1$: I want my muscles to contract, and I believe that contracting my muscles will result in my muscles contracting), and a primary reason for clenching my fist ($PR_2$: I want my muscles to contract, and I believe that clenching my fist will result in my muscles contracting), as I know this will lead to my muscles contracting in the appropriate way. If I, in fact, contract my muscles by clenching my fist, then it seems that both primary reasons could be considered to be causes of the action, and hence, as the first basic movement is the contracting of the muscles, then this is the movement I must be considered to have brought about directly. $PR_1$ might be thought to cause the action, as if I did not have the belief it includes, then I would not have any reason to consider the means by which I could accomplish the contracting of my muscles. This, though suggests a possible solution: that the primary reason should contain some reference to the means by which the action should come about, specifically, that the action should be performed directly. Thus the primary reason should include a desire for end, *E*, and a belief that directly *BA*-ing would result in *E*. So, the agent may have a reason with the content of $PR_1$, but which I now claim is not in fact a

primary reason, and this might cause the basic movement, along with primary reason $PR_2^*$: I want my muscles to contract, and I believe that directly clenching my fist will result in my muscles contracting. I might also have a primary reason, $PR_1^*$: I want my muscles to contract, and I believe that directly contracting my muscles will result in my muscles contracting. Such a primary reason, though, is not among the causes of the clenching of my fist or the contracting of my muscles.

I have suggested that primary reasons are the direct causes of basic movements, but primary reasons are not events, and are therefore not candidates for causes of events. Davidson suggested that it might be my coming to have a primary reason that is the event that causes an action. But it seems that I might have the reason long before I actually act on it. The event that triggers the action might be a change in some other psychological state of mine, such as a change in weight attached to another reason. A full discussion of the issues raised by this point would require a consideration of causation that lies outside the scope of this thesis, but I suspect that introducing a time component into the primary reason might be the answer.[15] So again, while the agent might have reasons with the content of the modified $PR_1$ and $PR_2$, these should no longer be considered to be primary reasons. Instead, perhaps, primary reasons could be thought of as comprising a desire for *E* now (next, in an hour or as soon as possible), and a belief that *BA*-ing now (or next[16]) would result in *E* now (next, in an hour or as soon as possible). Such primary reasons might be thought to arise when *E* became the most urgent goal for the agent, and their onset to cause basic movements.

I think that the sort of directness constraint I have suggested should be workable, though it might well need refining. The principle objection to it that I can see is that we simply do not have any psychological states

[15] Thalberg (1984) has discussed how adding a time component to intention could help solve the deviant causal chain problem, and my idea is inspired by this.

[16] c.f. Velleman (1989, p. 142), where he argues that we need not have a reason to act now, that next can be enough.

like the beliefs and desires that I have argued constitute primary reasons. I agree that we are not usually aware of any such states, though I find it quite plausible that we have them. I hope to address the problem that this raises in the next section, so, for now, I will take myself to have gone some way to satisfying condition (3), providing a directness constraint, though I do not pretend my position is wholly satisfactory.

### *5.3. Agents and actions: expanding the causal story*

The theory I have sketched so far is as lacking an account of the agent in action as the standard story of human action as the account proposed by Davidson. I have, after serious consideration, rejected the idea that describing actions as tryings is the appropriate way to introduce the agent into a causal theory of action, and I will now have to give an alternative account of how the agent might be involved in his actions. My theory so far claims that an action occurs when a primary reason causes a basic movement. There is nothing here that essentially involves the agent, other than as the bearer of the psychological states that make up the primary reason. This is some sort of involvement, but, as Velleman's example of alienating a friend as the result of unconscious desires and Frankfurt's example of the unwilling drug addict show, not nearly enough. This is consistent with my view of action as applying in a wide variety of cases where the agent is not involved to the greatest extent possible. However, I think that the story so far is enough to combat internal deviant causal chain examples, and so it seems that Chisholm was mistaken that the primary force of these examples was to show up the absence of the agent from the standard causal story. This is not, however, necessarily the case. I am happy to use the term 'action' to apply to many things the agent does, with only minimal involvement from him (as the bearer of the psychological states that make up the primary reason). I recognise that others may not be comfortable with this very weak concept of action. In showing how the agent can be involved in his

actions, then, I might also be seen as showing how a stronger concept of action is possible.

Adding more to either the primary reason or the basic movement would disturb the balance of the account so far. Therefore, in order to satisfy condition (4) above, that is, to give a plausible rôle for the agent in his actions, and at the same time construct a theory that is causal, thereby satisfying condition (1), it seems that we must look to the causes of either primary reasons or basic movements. David Velleman (1989, 1992a) gives an account that is intended to show how the agent can be involved in his actions *via* the causal antecedents of his bodily movements, and his ideas, therefore, seem excellent material for discussion.

Velleman argues that, if reasons are to act as reasons, and not merely as causes, then they must cause actions *via* the agent's grasp of their significance. He must act on them because he understands them: they must not be enough to cause a bodily movement without his understanding, and therefore involvement. He says, 'the influence appropriate to a recommendation, of any kind, must involve the agent's grasp of its meaning' (1989, p. 191). To illustrate, he points out that, if someone says "Jump" and I jump, I could have jumped because it startled me, or because I understood and did what they said. The theory I have sketched so far would exclude the former from being an action, as I would not in this case have a primary reason for jumping, or, even if I did, it would not be the cause, let alone the direct cause, of the jump. Velleman's thought so far does not seem to hold out any prospect of showing what is wrong with Frankfurt's example of the unwilling drug addict: though she does want to give up drugs, it seems fair to say that she would not have performed the action of obtaining drugs unless she understood her craving as a craving for drugs, and so her reason as a reason for obtaining drugs. However, in the example given by Velleman, where he alienates a friend as a result of unconscious reasons, it seems that the agent is not aware of the significance of his reason: he is not even aware that he has

it. Thus Velleman's idea that reasons should cause actions *via* some grasp of their significance does illuminate something that is missing, and he argues that, whenever reasons do cause actions in this way, then we would say that the agent is the author of his actions: that his actions being caused in this way amount to his being the author of them.

Having established something that is lacking from the standard causal story, Velleman proceeds to show how a causal account can remedy the situation. He remarks that some thinkers have argued that, if a reason is to act as a reason, not merely a cause, then it must be insufficient itself to cause an action. This thought is sometimes expressed as the idea that reasons must have normative force, as well as simply causal power. But, Velleman argues, this does not entail that there is anything other than common-or-garden event causation going on when we act. The reason alone, perhaps, must not be sufficient, but the remainder of the causal power needed to effect the action could be provided by desires of the agent other than that composing the primary reason. Velleman then turns his attention to finding desires whose involvement in causing an action would amount to the agent being the author of that action. In 'What Happens When Someone Acts?' (1992a), he suggests that a desire to act for reasons could play this rôle. When an agent recognises that a belief-desire pair constitute a reason for him to act, then this desire throws its causal weight behind that reason. He argues that the causal involvement in an action of the desire to act for reasons amounts to the agent's authorship of that action. It seems that we do have such a desire: we do generally want to act for reasons, rather than be swept along on a tide of events, and, as Velleman suggests, to repudiate this desire seems to be to repudiate our rôle as agents. Velleman argues that the desire to act for reasons cannot operate without the agent's awareness, but I am not entirely sure this is the case. However, if it is *recognition* that a certain belief-desire pair constitutes a reason, and that acting on it would satisfy the desire to act for reasons, that must play a causal rôle, then the possibility of unconscious operation is eliminated. That this is the sort of

idea Velleman has is clearer in *Practical Reflection* (1989), than in the paper I have been discussing. Also in his book, he makes some suggestions that, if accepted, would give us the resources to explain the partial absence of the agent in Frankfurt's drug addict example (p. 25, above). Here, he argues that there are a number of desires that could be considered to play the rôle of the agent in action, not simply the desire to act for reasons. We might, for example, want to act in such a way as to promote long term aims, or to live a life that we view as desirable. We might want, not only to act for reasons, but to act for *good* reasons (Velleman himself argues that this is not *necessary* for an agent's involvement in his actions (1992b), but perhaps it is necessary for the most full-blooded of full-blooded actions). Thus we have a desire to act from moment to moment in accordance with reasons, rather than be swept along by events, but we may have a number of further desires that involve more long term aims. It seems plausible to me that some such long term aims should be seen as partly constitutive of the agent, that our long term goals and ideals are part of who we are, and that, if recognition of the impact of acting on a reason on these central goals is not involved in causing an action of mine, then, whilst I still might be the author of the action, I am not involved in it to the fullest extent possible. What is more, if I recognise that the action proposed by a reason does *not* further these central goals, but I go ahead and perform it anyway, then it seems obvious that I am not wholly involved in the action. Thus, if recognition that action $A$ would frustrate the achievement of certain goals important to me is not enough to cause some alternative action $B$ (or no action at all), then we would probably say that the agent was not fully involved in his actions. This seems to be the case in Frankfurt's example: the agent has the goal of kicking her habit, and recognises that obtaining and taking drugs would frustrate her goal, but the desire for drugs (which is supported in this case by the desire to act for reasons) is strong enough to cause her action anyway.

I agree with Velleman that introducing other desires of the agent, desires that are central to the agent, into the causal story is the right way to go about introducing the agent himself into the causal story. Particularly interesting are the desires for self-awareness and self-understanding (1989) that he suggests must play a causal rôle in an agent's action: I have said that for intentional action the agent should know both what he is doing and why. However, despite Velleman's apparent identification of the problems posed by his and Frankfurt's examples (1992a) and those raised by deviant causal chain counterexamples, it seems that his account is still vulnerable to the latter. To demonstrate this, I will expand Davidson's climber example. The climber has a reason for dropping his companion, and we must suppose that the agent also grasped the significance of this reason, that is, he recognised that acting on it would satisfy his desires for reasons, as, if he did not, then he would not have been shocked and so dropped the rope. He may even have had some longer term aim, such as surviving the climb, that he recognised would be served by dropping his companion. Presumably, there were also desires important to him that the action proposed by his reason for dropping his companion conflicted with, else he probably would not have been shocked, but, nonetheless, Velleman's account cannot explain why his dropping his companion is not an action, let alone why the agent is not at all involved with his action. It could be that the recognition that the reason for dropping his companion is a reason and recognition that dropping his companion would promote goals important to him that cause his shock and his hand to open. A directness constraint is still required, in order to escape deviant causal chain counterexamples.

A further problem with Velleman's account is that he does not look closely at the plausibility of the claim that reasons cause actions. Consideration of Hornsby's 'bodily movement$_T$' *vs.* 'bodily movement$_I$' distinction and the necessity of identifying basic actions have taken me some way from this simple explanation of the relationship between

reasons and causes. On my account, reasons for action, plus other psychological events in the agent, cause what I have called 'primary reasons' (which are not quite the same as Davidson's primary reasons), which then directly cause basic movements. I have said that an action occurs when a primary reason causes a basic movement, and so the statement 'reasons cause actions' is not completely wrong on my account. However, it is not a statement that can be made strictly within my general theory of actions, as action is something that emerges from the theory and not an element within it. It also seems to be false that we need to recognise anything like the things I have called primary reasons as reasons for us to be fully involved in our action. Primary reasons as I have defined them are reasons for controlling our bodies, and involve beliefs about moving our bodies. The plausibility of Velleman's account relies on the fact that the sorts of reasons he mentions are ones that are likely to figure in practical reasoning, it does not seem that we are often aware of beliefs that moving our bodies in a certain way would achieve a desired end, or that recognition that we have such beliefs must play any part in our acting.

Thus, while Velleman argues that the desire to act for reasons, for example, and the reason for acting together cause the action, I must find a way to cash this in in a manner compatible with my theory. I suggest that the reason for acting, and other desires central to the agent (plus beliefs about the time, perhaps) cause the primary reason, which then causes the basic movement. The recognition that a particular action would serve the ends of a desire to act for reasons or another central aim of the agent could cause the primary reason, along with a reason of similar form to, but more general than, the primary reason. For example, I might have a desire to turn on the light now, and believe that flipping the switch is a means to this end. I might also believe that moving my body in a certain way would result in the switch being flipped. My recognition that flipping the switch would satisfy my desire to act for reasons might then cause, along with the desire to turn on the light and

the other beliefs of mine cause a primary reason for turning on the light now, constituting a desire to turn on the light now, and a belief that moving my body now in a certain way (moving to the switch and then moving my hand) will accomplish this aim. This might be thought to allow for another sort of deviancy that would result in an agent's lesser involvement with the action, namely a strange causal chain from reason to primary reason (imagine the reason and the desire to act for reasons somehow caused a primary reason to hop around like a frog). This sort of deviancy would show itself as irrationality in the agent, and I agree that irrationality, whilst not completely removing an agent from his action, does reduce his involvement. We might, at this point, specify that the agent is only fully involved if the causal chain from reasons and desires of the agent to primary reasons roughly follows a pattern of valid inference appropriate for practical reasoning. The agent need not be aware of all the steps in order for the reasoning to be valid in this way.

In this way, Velleman's account can give us the resources to meet condition (4), to give a plausible rôle for the agent in his action. However, it occurs to me that the account given so far does not exhaust the ways in which the agent can be involved in his actions. I am not sure that an agent needs to be aware of the reasons for which he is acting in order to be involved, though I do think this is necessary if the action is to be intentional or if the agent is to be fully involved. Consider Velleman's friend example. If the desire to break up with his friend stood alone, or along only with a perhaps temporary, unconscious, dislike of his friend, then I do not think we would have much temptation to say he was involved in the action. However, if he had unconsciously believed that keeping this friend would be a bar to his long-term happiness or some other goal that was very important to him, and realised this too when considering his behaviour after the fact, then we might want to say that he was involved in the action. Thus, it might be that the agent can be involved to some extent by the causal operation of his central goals, even if a recognition that an action would promote these goals is not causally

involved, and even in the absence of the recognition that he has a reason for the action he performs. Similar considerations apply for Frankfurt's drug addict example. If we were particularly unsympathetic, we might think that the agent was 'showing her true colours' when she acted on her impulse to obtain drugs: once a junkie, always a junkie, perhaps. That is, we might think that her addiction was such a defining feature of her existence, that any action based upon it is *her* action, in a way that actions based on her attempts to reform are not. I could also envisage an argument in which facts were brought to bear on whether this was a correct interpretation of events ("No, she really has come to recognise the harm her addiction is causing, and has repudiated the part of her that wants to take drugs. She sincerely wants to reform."). Thus, I suggest that an agent need not be aware of his reasons for acting and their relation to his desire to act for reasons or other central desires of his in order to be involved in his action. He will be involved to some extent even if one or some of his central desires plays a rôle in causing his primary reason without his awareness of its rôle.

## *5.4. Degrees of action and agency*

The way in which I have tried to give a plausible rôle to the agent in his actions, and thereby meet condition (4) above, suggests how I might also meet condition (5). That is, I now have the resources to show how a general theory of action might reflect our uncertainty in some cases regarding whether an action has occurred, and the different extents to which we think that an agent can be involved in his actions.

I have suggested that, in Chisholm's example, where a nephew runs over his uncle as a result of his bad driving, we might be unsure whether the fact that he was driving intentionally conferred actionhood on his hitting the pedestrian. I have argued that primary reasons cause basic movements, but other things can also cause similar movements: shock or horror, for example. I think that something is certainly an action when a primary reason, perhaps plus other reasons and psychological states of

the agent that involve him in his action, is enough to cause the action. However, if these are not enough to cause the movement, and something other than a reason or a desire central to him is required, such as shock, then our confidence that what occurred was indeed an action is reduced. Thus, the extent to which a certain event should be considered an action is affected by the extent to which it is enough for the occurrence of that event that reasons of the agent, and central desires of the agent caused a primary reason which caused a basic movement. If other factors were required for the occurrence of the disputed action event, then we can say only that it has important similarities with actions, but that it is not fully an action. If the primary reason directly causing the basic movement were not required at all for the occurrence of the event, then no action at all has taken place.

That an agent's involvement in his action can be a matter of degree is entailed by the way I have answered the question of how he is involved in action at all. I have argued that an agent might be seen as author of his actions in virtue of his recognition that he has certain reasons and that acting on them would contribute to his desired end of acting on reasons playing a part in causing a primary reason that directly causes the appropriate basic movement. He might be still further involved by the causal involvement of a recognition that the action under consideration promotes some further goal he holds dear. He could also be involved, though not fully, by the causal influence of some central aim of his on his behaviour without his recognition of its influence. Further, if a recognition that an action would frustrate the achievement of some important goal occurs, but the action occurs anyway, then the agent's involvement is reduced. This means that we must not only look to the causes of the action in order to determine how much the agent was involved, but must also look to events that occurred that perhaps did not cause the action. If an action is contrary to some central aim of the agent, even if he does not realise it, then we might think he was not fully involved in it (though if the contrariness were sufficiently subtle, we

might not think it has any effect on the agent's involvement). The latter two concerns that count against the involvement of the agent are not necessarily, I think, analysable in terms of the causation of actions. Thus an agent's involvement in his actions is not simply a matter of what caused his actions. Though this is not a wholly desirable result for my purposes of constructing a causal theory of action, I do not think that it is particularly damaging. There are no unexplained concepts, at any rate, and nothing incompatible with my theory. It does seem, though, that if the agent is to be fully involved in his action that we must add the constraints that he did not recognise the action to be incompatible with some central goal of his, nor was it in fact incompatible in a way that it would be reasonable to expect the agent to have noticed, had he considered it. In order for an agent to be fully involved in his action, all these factors should pull together. The agent should, first, recognise his reason as a reason for an action that, if performed, would satisfy the desire to act for reasons, second, if the action does promote some long term aim, he should recognise this, and these first two events of recognition should causally contribute to the formation of a primary reason. Third, the agent should not have recognised that the proposed action was incompatible with or detrimental to the achievement of any of his important long term aims, and nor should the action have been so incompatible or detrimental in a way that it would be reasonable to expect the agent to have noticed had he thought about it. However, it seems unlikely that all these conditions would ever be satisfied. It is possible, and perhaps even likely, that an agent has conflicting central goals or desires. Frankfurt's drug addict, we might think, has both the desire for drugs as a result of her addiction, and the desire to break her habit, to which latter she might attach great importance. We might say that she could not be fully involved in her actions until she overcame her addiction, even if she no longer acted on it. We might also attach varying weight to desires, all of which we nonetheless see as central to us. This might again lead to degrees of involvement: more if the desire causing the

action is one which is more important than a desire which is flouted, and less if the weights are reversed.

The theory that I have been proposing, then, reflects the confusion that we might sometimes have about the agent's involvement in his actions. The best answer it can give, like the best answer we can ordinarily give, is to point out ways in which the agent is involved in the action, and ways in which he is not.

## *5.5. Conclusion*

In this thesis, I think I have made significant progress towards providing a general theory of action that satisfies the five conditions set out in §2.10. An action, I have argued, is what happens when a primary reason directly causes a basic movement. The agent is involved in his action to the degree that the primary reason that directly causes the basic movement is caused by the agents reasons and by recognition that acting in this way would promote his end of acting for reasons and/or other central goals of his, and that the action is caused by desires central to him, even in the absence of his recognition that the action promotes their satisfaction. I have also discussed how the theory accounts for the intuitively plausible idea that action and agency are matters of degree.

To conclude, I will examine how well the theory I have sketched will accommodate the intuitive beliefs we have that action explanation is causal explanation that motivated the search for a causal theory of action in the first place. I will consider three examples. First, that reasons for actions cause actions, for example, my reason for turning on the light caused my turning on of the light. I have argued that, in an ordinary case of acting for reasons, my reason for turning on the light (plus other beliefs and desires of mine) causes a primary reason for moving my body in such a manner as to turn on the light, which directly causes my body to move in the appropriate manner. The direct causation of the basic movement by the primary reason amounts to an action being performed, though the

action cannot be identified with the basic movement (or, indeed, the primary reason). Thus reasons can be said to cause actions, in a sense, but reasons are elements in the theory whereas actions emerge from it, so the statement 'reasons cause actions' is not exactly right from the point of view of the theory. Second, we might say that agents cause actions, for example, I caused my turning on of the light. The causal story here is as in the first example, and we might say that my reasons, and my recognition that acting on this reason would promote my aim of acting for reasons, represented me in this action, they serve, as Velleman suggested, as a functional equivalent for me, and their causing the primary reason that directly causes the basic movement, amounts to me causing the action. Again, though, neither actions nor events of agent-causation are elements in the theory, they emerge in cases where the theory describes events. Third, we might say that agents cause events that are not actions, for example, I caused the light to turn on. Insofar as this is true, it is true in virtue of the fact that my reasons, and my recognition that acting on this reason would promote my aim of acting for reasons, caused a primary reason which directly caused a basic movement which in turn caused the turning on of the light.

The theory I have sketched, then, explains the intuitive force of statements such as 'reasons cause actions', 'agents cause actions' and 'agents cause physical events', but also suggests why, on closer examination, we are unsatisfied that they are really true. This is because neither events of agent-causation, nor actions themselves, appear in the true causal-explanatory statements that underpin the relationship between an agent and his actions.

## Bibliography


Annas, Julia (1978) 'How Basic are Basic Actions?' *Proceedings of the Aristotelian Society LXXXVII*, pp. 195-213.

Anscombe, G. E. M. (1957) *Intention.* Oxford: Blackwell.

Baier, Annette (1971) 'The Search for Basic Actions'. *American Philosophical Quarterly VIII*, pp. 161-170.

Brand, Myles (1968) 'Danto on Basic Actions' in *Noûs* (1968) pp. 187-190.

Brand, Myles (1984) *Intending and Acting: Towards a Naturalized Action Theory.* Cambridge, Mass.: M. I. T. Press.

Chisholm (1966) 'Freedom and Action' , in Lehrer (1966) pp. 11-44.

Churchland, P. M. (1970) 'The Logical Character of Action Explanations', in *Philosophical Review* 79 (1970) pp. 214-236.

Danto, A. (1965) 'Basic Actions', in *American Philosophical Quarterly* (1965) pp. 141-148.

Davidson, Donald (1963) 'Actions, Reasons, and Causes', reprinted in Davidson (1980) pp. 3-19.

Davidson, Donald (1971) 'Agency', reprinted in Davidson (1980) pp. 43-61.

Davidson, Donald (1973) 'Freedom to Act', reprinted in Davidson (1980) pp. 63-81.

Davidson, Donald (1980) *Essays on Actions and Events.* Oxford: Clarendon Press.

Frankfurt, Harry (1988) *The Importance of What We Care About.* Cambridge: Cambridge University Press.

Goldman, Alvin I. (1970) *A Theory if Human Action.* Englewood Cliffs, N. J.: Prentice Hall.

Hornsby, Jennifer (1980) *Actions.* London: Routledge & Kegan Paul.

Hornsby, Jennifer (1993) 'Agency and Causal Explanation', reprinted in Mele (1997, pp. 283-307).

Hornsby, Jennifer (forthcoming) 'The Poverty of Action Theory' in *Philosophical Inquiry: International Quarterly.*

James, William (1890) *Principles of Psychology*, authorised edition in two unabridged volumes. New York: Dover, 1950.

Lehrer, K. ed. (1966) *Freedom and Determinism.* New York: Random House.

Mele, Alfred R. (1997) *The Philosophy of Action.* Oxford: Oxford University Press.

Moya, Carlos (1990) *The Philosophy of Action.* Cambridge: Polity Press.

O'Shaughnessy, Brian (1973) 'Trying (as the mental "pineal gland")' in *The Journal of Philosophy* Volume LXX, No 13 pp. 365-386.

Pink, Thomas (1991) 'Purposive Intending', in *Mind* Vol. 100 (1991, pp. 343-359).

Pink, Thomas (1996) *The Psychology of Freedom.* Cambridge: Cambridge University Press.

Pink, Thomas (1997) 'Reason and Agency' in *Proceedings of the Aristotelian Society* (1997) pp. 263-280.

Thalberg, Irving J. (1977) *Perception, Emotion and Action: A Component Approach.* Oxford: Blackwell.

Thalberg, Irving J. (1984) 'Do Our Intentions Cause Our Intentional Actions?' in *American Philosophical Quarterly*, 21 (1984) pp. 249-260.

Velleman, J. David (1989) Practical Reflection. Princeton: Princeton University Press.

Velleman, J. David (1992a) 'What Happens When Someone Acts?' in *Mind* (1992) pp. 461-481.

Velleman, J. David (1992b) 'The Guise of the Good' in *Noûs* 26, pp. 3-26.

Wilkins, John (1668) *An Essay Towards a Real Character, and a Philosophical Language.* London: printed for Sa. Gellibrand.

Wilson, George (1989) *The Intentionality of Human Action.* Stanford: Stanford University Press.

Wright, G. H. von (1971) *Explanation and Understanding.* London: Routledge & Kegan Paul.