

DESIRES AS DETERMINANTS OF

ACTION

Ian Hilton Dearden



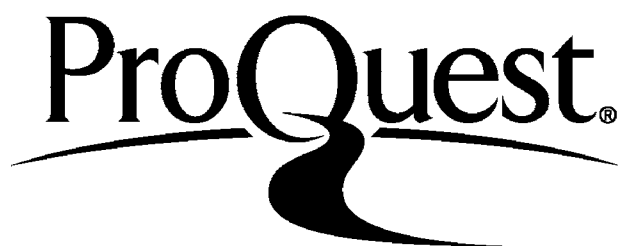
ProQuest Number: 10098252

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10098252

Published by ProQuest LLC(2016). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code.  
Microform Edition © ProQuest LLC.

ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346

## CONTENTS

	page
Abstract	i
Chapter 1 Action as Criterially Linked to Desire	1
Chapter 2 Logical and Causal Relations between Action and Desire	14
Chapter 3 The Various Ways of Explaining Actions	21
Chapter 4 Actions, Reactions and Bodily Movements	43
Chapter 5 Intentions: For the Future	57
Chapter 6 Intentions: as Involved in Action	72
Chapter 7 Mental Action	80
Chapter 8 The Explanation, Understanding and Criticism of Action	107
Footnotes	134
Bibliography	159

Total number of pages = 163

Abstract

A supposed logical connection between an unimpeded desire and action, which might be thought to preclude a causal connection, is criticised (chapter 1). The thesis that if a desire is to explain an action as a reason for that action, then the desire must match that action in a certain way is accepted, but it is argued that this is not an objection to a causal theory of action (chapter 2). It is maintained that (i) explanations of action are explanations in terms of the agent's reasons (ii) there may be reasons for acting other than desires but these motivate in a way to be likened to the way in which desires motivate (iii) a causal force must be given to the "because" implied in the statement of the reason why someone acted (chapter 3). An attempt is made to distinguish actions motivated by desire or fear from bodily reactions characteristic thereof. Certain actions for which one has no reason are considered (chapter 4). An attempt is made to analyse intentions to do something in the future in terms of desire and belief, but this seems reductive (chapter 5). However, this does not vitiate the previous analysis of action (chapter 6). The subject of mental action is broached. It is suggested that the most fruitful approach involves considering the limitations of mental action: the only clear cases uncovered involve the direction of one's attention (chapter 7). One's understanding of another's action is considered. It is maintained that an explanation in terms of his reasons has its own kind of completeness, but such a complete explanation would not be deemed adequate for all purposes. The attitudes one takes up to another because of his actions are discussed and while it is admitted that such attitudes could not simply be abandoned, there remain problems about the justification of them and actions motivated by them (chapter 8).



## Chapter 1

### Action as Criterially Linked to Desire

It is clear that there is a close connection between wanting and acting. "The primitive sign of wanting is trying to get" writes Miss Anscombe. (1) But how close is this connection? One might be led to stress the conceptual relation between desire and action by the need to avoid seeing desiring as an inward process the concept of which we acquire by observing it in ourselves. One might then take the actions that people perform as a way of telling what their desires are. (This procedure has the merit of being applicable to animals - unlike, for example, treating avowals of desire as the criterion of desire). It is difficult to see how, on this basis, one could allow the possibility of someone's having a desire on which he never acts, nor on this simple view, which does not involve the concept of belief, could one introduce the concept of actions which are a means to an end, that is, how one could infer a desire for an end other than the action itself from an action; the criterion allows us to infer from an action A a desire to perform A but how does it allow us to infer a desire for something that is the goal of the action? Clearly, the action-criterion cannot be used in isolation - in fact, as a criterion it seems to be circular, for in order to distinguish actions from mere bodily movements one must know or at least assume something about the desires of the putative agent. And if it is said that it is the intentions rather than the desires of the putative agent that we must consider, then one can reply that the same difficulties will arise if one tries to infer the intentions of a putative agent from a piece of behaviour taken in isolation. (I am using 'behaviour' to cover both actions and mere bodily movements). There seems to be no way of using an agent's behaviour alone, as yet uninterpreted as an action or otherwise, in order to establish either his desires or intentions.

A full consideration of the evidence we use for determining an agent's desires would make reference to avowals and our knowledge of the needs of the agent and of his beliefs. In the case of animals, we must place needs at the centre of the stage because of the absence of verbal avowals on their part. But needs are important in

identifying and, in particular, in finding some intelligibility in the desires of human beings. Beliefs, of course, are themselves ascribed on the basis of criteria which are in the broadest sense behavioural, and furthermore beliefs can often be ascribed to an agent on the basis of its behaviour because we know its desires, rather than vice versa. To show just how it is possible that our ascriptions of belief and desire should ever get off the ground, and at what point, would be a task of some complexity and I shall not attempt it here (needs would obviously play a fundamental role as they can be ascribed to an organism on the basis of its being an organism of a certain sort or an organism at all, as would perception, though both needs and perceptions are ascribed to organisms or classes of organisms only in as far as we are able to ascribe actions to them). One other form of evidence we have for an agent's desires comes from the fact that certain desires are associated with bodily reactions (not actions) - a non-contr<sup>o</sup>versial example would be sexual desire, whereas a somewhat more tendentious example might be fear, for I shall argue that fear is to be bracketted rather than contrasted with desire.

Having said that it is not possible to take action as the sole criterion of desire, I want to investigate the possibility and consequences of taking it as, in a certain sense, an overriding criterion, or more precisely of the view that the absence of action in certain circumstances shows that someone does not have a certain desire. This view can be put as follows: if someone claims (sincerely or otherwise) to desire A, then if he has no countervailing desire, and believes that it is possible for him to satisfy his desire for A, then if he does not act in order to satisfy the supposed desire, he does not really desire A. I shall call this the thesis of action-criterion primacy ("overridingness" would be better, but ugly). If we assume that the person was sincere in his avowal, we face the problem of characterizing the entity which masqueraded as a desire, and indeed of putting a name to it. One reply might be: "There is no such entity. Ordinary language gives us a lead. 'He thought he desired A' and 'It seemed to him that he desired A' are the appropriate idioms and the latter no more implies the existence of a seeming-desire than 'It seemed to him that he saw an apple' entails 'He saw a seeming-apple'. We can recognise the phenomenon of people wrongly believing that they desire something without postulating anything more than the false belief."

This reply does not seem adequate to the case of someone seeming to have a desire of some intensity. "He falsely believed that he longed or yearned for A" seems an insufficient characterization of someone who has the experience of having a powerful desire for A but does not do anything about it though he believes he can, and has no countervailing desires. One wants to say that at least he must have something like a desire of which adjectives of intensity could be predicated. If the action-criterion does have the primacy in question, then the locution "I believe I desire A" is never illegitimate, at least when one is not actually engaged in action devoted to satisfying the desire, though it might seem pragmatically out of place. Tentativeness will always have a certain justification. To avow a desire will be to commit oneself to a categorical or conditional prediction. If the implicit prediction is not fulfilled, one will have to withdraw one's claim to have desired A, or at least admit that one no longer desires A. If one takes the former course, one will have to admit that this discovery is not prompted by any discovery about oneself at the time one made the avowal, but by a discovery about oneself at a later time. So there need be no present means of deciding whether I really do desire A or merely think I do. But against this ~~are~~ we not inclined to say that we know what we want in a way that is not corrigible simply by our failure to act in a certain situation? Or rather that we do not have a way of knowing what we desire, nor a way of vetting or discounting the deliverances of any such avenue of knowledge? It does not seem that we allow for the constant possibility of such radical mistakes concerning our own present desires.

If someone claims to desire A, and for no assignable reason fails to perform actions directed towards satisfying the desire for A when he believes he has the opportunity to do so, then so long as we regard him as sincere, our actual practice is not to deny that he desires A but, to assume that there must be some factors unknown to us which can account for the failure. The acceptance of this point is of course consonant with adherence to the primacy of the action-criterion. Since Freud, unconscious countervailing desires would be considered particularly promising candidates. It ought to be mentioned at this juncture that I am not maintaining that one cannot be mistaken about one's desires; it would be inconsistent of me to maintain this as to allow unconscious desires is to allow that one can have desires that one does not believe one has and in some cases believes one does not have. What

seems to be true is that the move of denying that someone has a desire which he believes or avows he has, is one that is not often made.

Even if there is no reason for the failure to act which can be counted as a countervailing desire, or absence of relevant beliefs about the possibility of satisfying the desire or some impediment such as paralysis, is there any reason why the desire should not simply fail to lead to action? An answer to this needs some deployment of extra considerations; it is not sufficient to say that the prior desire is distinguishable from the action so that the former might occur without the latter for this begs the very question at issue. The action-criterion primacy (ACP) theorist would say that when there is no action, there is no desire. However, everyone would admit that there are certain times when a person is entitled to say, "I desire A", although on the view of the ACP theorist he might be wrong. And these states of affairs are certainly separable from any later action to which they might lead. (To the ACP theorist the class of such states of affairs would be disjoint).

Might it be a matter for stipulation whether we call the states of a person in such cases "desires" or "the states in which he can sincerely say 'I desire'..." ". Scientific observations could make one convention seem the more reasonable. If no great difference were found between the states in which people sincerely say 'I desire A' and this is followed by action and those in which it is not (even though the agent believes he can satisfy the desire and he has no countervailing desires) then the practice of designating all such states "desires" would seem validated. It may seem odd to consider this issue as one of stipulation, when I am making a claim as to what our actual usage is; for one contrasts "stipulative" with "<sup>re</sup>portive" definitions. But the claim of the ACP theorist can be seen as a stipulation in conflict with our actual practice - and one could imagine scientific discoveries that would give it point.

The idea that the claim of the ACP theorist is really a stipulation is supported by the difficulty of finding features of our practice of ascribing desires which reflect it. Had, per impossibile, action been our sole criterion of desire, then it would have been natural to allow it to override mere symptoms or correlates. In order

to avoid unverifiable hypotheses one might make action a necessary and sufficient condition for the ascription of a desire. And if it were possible to use the action-criterion on its own this would confer upon it a certain independence which would allow us to introduce a new concept related to, but distinct from, that of desire, which made use solely of the action-criterion. But, in fact, the action-criterion is shackled to those based on avowals and needs and so the procedure for applying the action-criterion primacy thesis will have to take the form of using other evidence first and then demanding that the action-criterion be satisfied. It would be a necessary condition that the action-criterion be satisfied, but only a sufficient condition in that the other criteria would have to be satisfied, or assumed satisfied ~~and~~ before it could be applied. The primacy of the action-criterion would seem easily detachable from the notion of desire, even if it were part of our actual practice of ascribing desires. Of course, the fact that action can be evidence for desire is a conceptual truth and this is not detachable without radical alterations to our concept of desire, but it would be otherwise with the requirement of action. I do not, of course, wish to maintain that it is coherent to suppose that there might be a race of people who desired but never acted upon their desires.

The ACP thesis, in one form at least, as we have seen, seems counter-intuitive as regards what seem to be strong desires, but which do not lead to action in the relevant circumstances. One feels compelled to postulate something masquerading as a desire at least. Suppose the putative desire actually led to bodily reactions (consider the desires involved in emotions, for example) that is suppose the person who sincerely claimed to have a certain desire, exhibited the bodily changes typical of the desire. It seems plainly wrong to <sup>say</sup> ~~deny~~ that the person does not have the desire merely because he does not act in the right circumstances. The ACP theorist has these options. He can regard bodily reactions as an alternative to action in his statement of the requirement for genuine desire. Or he can regard bodily reactions as rudimentary actions (a suggestion which, I shall suggest later, has a certain plausibility in some cases but not in all). Or ~~if~~ in the case of desires associated with emotions, he can divorce the emotion from the desire with which it is associated; he might say, for example, that although someone might be genuinely angry

with someone, he does not genuinely desire revenge, even if he thinks he does, if he takes no action in the appropriate circumstances. It seems to me that only the first of these options is plausible, and that it constitutes a weakening of the claim of the ACP theorist.

The ACP theorist would have to adopt the view that statements of the form "I desired A" when made by someone in the best possible position to do so are ~~always~~ inferential, not necessarily reached by conscious inference, but never resting merely on the memory of the supposed desire but also on his knowledge that he acted or is acting on that desire. It should not, however, be thought that the views of the ACP theorist present us with a reductive analysis of desiring. The other criteria are used but are subject to veto by the ACP thesis. However, this latter requirement leads to distinct peculiarities in the analysis of desire. For example, it is not clear that allowance is made for the possibility that someone might have a desire and lose it before an opportunity arises for him to satisfy it. It might be said that in such a case the person would know that he no longer had the desire before the opportunity arose. But why place such reliance on the man's awareness when one is so ready to override it in other cases? It is not obvious that someone could not have a desire which at some point he loses, but only comes to realise this when what would have been an opportunity to satisfy it arises. And how is the ACP theorist to distinguish this from his never having had the desire at all?

If someone has a desire and an opportunity to satisfy it presents itself, when does he act? Immediately? If not, and there are no considerations which would make a postponement of action desirable, when are we to say that he does not have the desire? - by the ACP thesis the time-lag cannot be extended to infinity. Any answer sounds like a stipulation, as does the requirement that the person must ~~act~~ immediately. Suppose the desire does not immediately lead to action, and whether or not it would have done, had there been more time available, the opportunity is lost. Either one abandons ACP in any form applicable to the present case. Or one <sup>denies</sup> ~~desires~~ the possibility and says that an unopposed desire must immediately lead to action as soon as one believes one can satisfy it. Or one affirms a counterfactual statement to the effect that the man would have acted if the opportunity had not been lost which is to adhere to

ACP but to cease using it criterially and to make it a logical principle, arguing not from inaction to lack of desire, but from desire to action. Finally, one could say that one just does not know whether the man had the desire. These difficulties can be avoided by not affirming ACP.

With some desires one never believes one has an opportunity to satisfy them, or they are never unopposed by considerations one believes weightier (perhaps the reason is that the desires are of short duration). Does one's readiness to regard them as genuine desires mean that one implicitly subscribes to various counterfactual and subjunctive conditional statements? Perhaps but some desires might always be decisively opposed by others given that one has the character one has: one has certain radically opposed elements in one's personality. Does this mean one is committed to subjunctive and conditional statements, where one believes that the conditional clause will never be true? Again perhaps, if one ever bothered to think about the matter, but these considerations seem to underline the limited applicability of the ACP requirement. It is hard to see the point of affirming a claim whose applicability is ~~as~~<sup>so</sup> restricted. In ~~as~~<sup>so</sup> many cases, one seems unable to use the ACP thesis criterially; one has to treat it as a logical truth and argue contrapositively from desire to an action which would occur if the circumstances were otherwise.

As far as I can see the main point of the ACP thesis or something like it is to uncover deception. If someone claims to want something, but in suitable circumstances takes no action this may create a strong presumption that he is a liar; but this only involves a weak, pragmatic form of the ACP thesis. If one has reason to believe in his sincerity, one does not have to assume that he is mistaken about having the desire. Indeed the strong form of the ACP thesis is less workable as a maxim for uncovering deception in that if, on the basis of his inaction in certain circumstances, one comes to the conclusion that someone's claim that he had a certain desire is false, there always remains the question as to whether he is mistaken. We can have independent evidence for a person's sincerity, for example his honesty at other times, which is more easily brought to bear on the occasion in question if we are unready to allow that the person is simply mistaken. For even a notorious liar can speak falsehoods simply

because he is mistaken.

There are still further difficulties in applying the ACP thesis. I have said that I do not wish to rule out the possibility of being mistaken about one's desires. One can be mistaken about which desire is moving one to act. When someone acts there is often the possibility that the desire he believes is moving him to act is <sup>not</sup> the only, or even the real desire moving him. So the application of ACP must always have an element of tentativeness. Someone says he wants A and does something he believes is a means of getting A, so his desire is validated - provided his action was really on the basis of the desire for A. Apart from the suspicion of circularity, it is still possible that he did not desire A because his action might not have been on the basis of a desire for A. If we apply ACP any scepticism we might entertain about a man's motivation is transferred to the question of the reality of certain of his desires. The ACP thesis makes our ascription of desires to people even more tentative than it seems at first sight. Not only is one's belief that one has a certain desire a hostage to fortune before one believes one has the opportunity to act; it still seems somewhat problematic afterwards.

Another peculiarity in the analysis of desire which the ACP thesis produces is that it seems to suggest the wrong counting-principle for periods during which one desires A. It suggests that if someone thinks he desires A over a certain stretch of time and during that time he repeatedly thinks he can satisfy the desire and has no countervailing desires etc. then if he did not act the first time, then he did not really desire A the first time. This leads us into the difficulties we encountered in considering whether someone must act immediately on a desire if he comes to believe he can satisfy it. But suppose the stretch of time were not continuous, that is there were periods when he did not believe he desired A or correctly believed he did not desire A. Then unless the supposed desire led to action in each of the time stretches when he believed he desired A, the number of periods during which he desired A must, to the ACP theorist, be less than the number of periods during which he thought he desired A.



So far, for simplicity, I have spoken of people who believe they desire something, when dealing with cases where the ACP requirement is not known to be satisfied. This idiom is suggested by the ACP requirement itself. But it would be better to speak of people who can sincerely avow that they have a certain desire; this does not imply that every time one has a conscious desire one must also have an accompanying belief that one has this desire. But even this idiom is not applicable to animals. It may be that one cannot construct cases of animals seeming to have a desire, but falling foul of the ACP thesis (what however of bodily reactions as evidence for desire?) and so one does not need a special idiom such as "apparently desiring" to describe them. However, the fact that the ACP theorist does need to resort to special idioms to describe cases, which I would prefer to call simply desires, does reveal a significant asymmetry in his treatment of desire.

This can be brought out as follows. The occasions when people do not desire certain things are of two types to the ACP theorist - (a) those where a person wrongly believes he desires A and (b) those where a person does not believe he desires A and does not desire A. Thus there is a lack of symmetry between real cases of desiring and spurious cases, (spurious because the action-criterion is not satisfied) - if one really has a desire one may or may not believe that one has the desire (animals presumably never do, unless one is going to make it a logical truth that desiring A entails believing one desires A) but in the spurious cases of desire, one must believe one has a genuine desire. I would suggest that even if there are spurious cases of desiring (one might admit them for other reasons than adherence to the ACP) the occurrence of such a case should not entail that the person has a false belief that he desires something. The sort of consideration that might lead one to allow the possibility of spurious cases of desiring, where no conscious insincerity is involved might exist in the following situation: a belief that one has a certain desire and that one acted because of it is needed to placate one's conscience and some unconscious mechanism generates a specious desire or a belief that one has such a desire. But whether the notion of a spurious desire is the right one to characterise such a case rather than say that of a false belief about the origin of the desire is not altogether clear. Nor indeed is the question as to whether such cases would be shown up

by the ACP thesis.

The ACP theorist might of course avoid involvement in the above difficulty by characterising apparent cases of desire which fall foul of the ACP thesis as say "pseudo-desires" and make no mention of false beliefs. The only trouble with this is that it seems that the difference between him and us seems to be merely terminological. We are back with the idea that the ACP thesis is simply a stipulation, and then the question arises as to what would give it point. Future scientific discoveries were suggested earlier.

However, the ACP theorist might feel that his convention already has point and does not simply await vindication by future discoveries. He might say that although we allow that a desire might not be followed by action even though there are no apparent impediments, beliefs to the effect that satisfaction of the desire is not possible or countervailing desires, we always assume that there must be some such explanation. Thus would be, in effect, to abandon the action-criterion and instead to claim a conceptual truth. And this, I think, could be admitted; whatever explanation, perhaps the neurophysiological level or perhaps some such factor as anxiety, could always be bracketted with impediments. We can afford to admit this attenuated version of the principle enshrined in the ACP thesis. It expresses our commitment to something like the principle of causality and a resolution to interpret whatever accounts for the failure of a desire to be followed by action in the relevant circumstances in a way that assimilates it to the considerations we normally regard as accounting for a person's failure to act on a desire. (2) If, however, the principle of universal causality is not true there may be no explanation of why a desire failed to be followed by action, but then there would be little point in refusing to call the supposed desire a desire when, *ex hypothesi*, it has no intrinsic features which explain the failure to act.

Even this version of the ACP thesis has its difficulties. It is true that when two similar entities produce different effects in similar circumstances, we expect that there must in fact be some relevant difference between them<sup>or</sup> in the circumstances. But suppose the failure of a desire to be followed by action in the right circumstances is explained by its possessing feature F. It

is not clear that this ( a property of the desire) can be regarded as an impediment to the natural operation of the desire. So it might be artificial to interpret the explanation of the failure of the desire to lead to action along lines which assimilate it to those explanations of failure to act we normally encounter. Of course, if the presence of F were a complete explanation for failure to act in the right circumstances, this might prompt a conceptual revision in which those desires with feature F might come to be no longer regarded as desires. And therefore certain seeming-desires which did not lead to action would no longer be regarded as desires. Thus an empirical discovery would make it reasonable to apply a criterion for genuine desires which would be similar to a strong version of ACP. And hence ACP might be used to demarcate the concept of desire in the same way as the presence or absence of F. But its use and point would depend on the generalization about F actually holding good.

In conclusion, therefore, it seems that the ACP thesis presents difficulties of formulation when it comes to characterising the apparent cases of desire which it deems spurious, and of application in that in many cases to be applied at all, it has to be applied contrapositively, so that one infers subjunctive or counterfactual statements about action from an assumption about desire. It derives its plausibility from two sources - firstly, from a confusion with a maxim for unmasking insincere avowals of desire, and secondly, from the fact that future discoveries could make its adoption more reasonable. This latter consideration, however, though it has more respectability, does not establish a logical connection between desire and action in a way that <sup>rules</sup> ~~makes~~ out a causal connection. For either it amounts to a resolution to interpret whatever explains a desire's failure to result in action in a way which is consonant with adherence to a certain conceptual claim or it amounts to a resolution to regard those apparent desires which do not lead to action as not really desires, because it is hoped that there will be some intrinsic feature of them which distinguishes them from real desires. If there is something which explains why an apparent desire does not lead to action this might be construed as an impediment to the actual operation of the desire (the first option) or as a feature which brands it as a spurious desire (the second option).

I have not quoted adherents of the ACP thesis because I have

not been able to find any that adopt it in quite the form in which I am considering it. Consider, for example, Raziel Abelson in his review of Richard Taylor's "Action and Purpose" (3) - "assume that Jones wants, intends, desires, or in some sense has a motive to open the window. What does this entail about what he will do? Well, it entails that he will open the window, but it does not entail this tout court. It entails that he will open the window provided that no reason arises for his not doing so (e.g. a hurricane is blowing outside) and provided nothing prevents him (e.g. he is not paralysed and the window is not stuck). The provisos here constitute the contextual limitation...on the entailment between motive and act. To say 'I want to open the window; nothing prevents me and I have no reason or motive not to, not even the motive of laziness, but still I won't open the window,' is senseless. What on earth could I mean by 'want'? "

It may well be that the avowal which Abelson rejects as senseless, is actually senseless, at least, if "I won't open the window" is construed as an expression of intention. But if we construe it in this way, can we avoid rejecting Jones' claim to have not motive or reason not to open the window? Might we not have to postulate at the very least a desire not to open the window, even if no further reason could be given for this desire? Or suppose Jones could have an intention not to open the window without anything like a desire underlying this. All Abelson's example could show is that it makes no sense to co<sup>n</sup>join an avowal of desire to do something, a claim that one has no countervailing reasons and that nothing prevents one, and an avowal of<sup>an</sup> intention not to do it. It shows nothing about whether a desire can fail to lead someone to act even when he has no countervailing reasons etc.

Can "I won't open the window" be construed as other than an intention? Possibly it can be taken as a mere prediction, perhaps based on inductive evidence. But if this is a possible interpretation, it is then not clear that Jones' utterance is senseless, unless one takes the line of assuming that there must be an explanation of Jones' failure to act if he does so fail and assuming that this can be bracketted either with countervailing reasons or with factors which prevent him from acting. And this move has already been dealt with. Nor is it clear that when Jones says "Nothing prevents me" he really means to exclude along with impediments

such as paralysis and the window being stuck, the possibility of various recondite states of affairs perhaps only describable in neurophysiological terms, of which he might have no notion. Abelson, it seems, wishes to endorse the ACP thesis, but his arguments seem only to reach as far as the much weaker conclusion discussed in the previous paragraph. (4) He wishes to adopt the ACP thesis, so as to be able to rule out a causal connection between desire and action; and as I have already hinted, my aim in refuting the thesis is to establish the possibility of such a connection. The idea that a simple entailment between desire and action would rule out a causal connection is familiar; in the next chapter I want to consider the view that another more abstruse logical connection would preclude such a relationship, and in doing so the idea of an opposition between logical and causal relationships will be brought more to the fore.

## Chapter 2

### Logical and Causal Relations between Action and Desire

I have argued that the ACP thesis is a dubious requirement to place on anything that is to count as a bona fide desire. The ACP thesis is a particular case of the more general claim that action and reason for action are logically connected. In this chapter I want to consider the claim that action and reason for action are logically connected because the descriptions offered of the action and reason must match. Consider the following argument:- "If someone performs action A because he desires D it may be the case that (a)  $D = \text{to perform A}$  or (b)  $D \neq \text{to perform A}$  but performing A is believed by the agent to be a means to getting D or a way of getting D. (By  $D = \text{to perform A}$  I mean that the description of D offered is 'to perform A'). In case (a) the descriptions of A and D already match, whereas in (b) we must assume that the desire for D is accompanied by a belief that performing A is a way of or a means to getting D, if the explanation of the action is to have any force. In both cases we have descriptions of action which in certain sense match a full description of the reason (in case (b) this could be brought out by constructing a practical syllogism - even if we allow that a desire for D together with the belief that performing A is a way or a means of getting D can explain the performance of action A without the mediation of a desire to perform A)".

The above argument can be strengthened further, but first it is necessary to explain what is usually the point behind the insistence that action and reason must match. If the description of the action and the reason must match if the reason is genuinely to explain the action, then the action and the reason are logically and not contingently connected, and hence, it is argued, the reason for the action cannot be its cause. (5) The argument might take the following form:- "In the simplest case of someone performing a bodily movement because he just wants to, there is a clear logical connection between his desiring to raise his arm and his raising his arm. If it is replied that there is no such logical connection between his desiring to raise his arm and his arm going up (and therefore the former can cause the latter, the whole process being somehow constitutive of the action) then we

can counter by pointing out that the logical distinctness of the desire and the bodily movement is illusory, as raising one's arm involves one's arm going up - thus the object of the desire involves the event it is supposed to cause. Furthermore, if the action of raising one's arm is analysed as the desire to raise one's arm causing one's arm to go up, then this looks circular as the definiendum seems to appear in the definiens. If the analysis is modified so that the raising of one's arm is taken as the desire that one's arm should go up causing one's arm to go up, then firstly we are back with explicit matching of the desire and its alleged effect and secondly we have not allowed for the efficacy of desires which are not only that a bodily movement should occur but that one should perform a bodily movement."

Now it is not claimed that a desire cannot cause a bodily movement. Presumably, a desire is not debarred from causing some totally irrelevant bodily movement. Perhaps even this could constitute an action provided that it is an action that is somehow bungled - for example, if I desire to raise my right arm and raise my left arm by mistake, though here the proponent of a logical connection might reply that a specification of the object of the desire (that I should raise my right arm) entails a less precise specification (that I should raise an arm) and that a description of the bodily movement (my left arm goes up) entails a less precise description (one of my arms goes up); and these less precise descriptions match. However if the logical connection theorist does take this line, he might end up by saying that a desire cannot cause anything as the object of the desire and its alleged effect might always be specifiable in terms sufficiently vague (say, "the occurrence of some event") as to make them seem logically connected. The idea that a desire cannot cause anything is deserving of consideration, but it seems more plausible to base it on the view that a desire is a state and not an event. (6)

In order to seem to show that a desire cannot be the cause of the bodily movement in the case of action, with full generality, the argument needs a further extension. For in the case where someone performs an action A because he desires D and believes that performing A is a means to satisfying the desire for D, it might be argued that the desire for D could cause the bodily movement involved in A, but that a desire to perform A to which it might lead could not ( and here it should be mentioned that the view

might be held that it is not so much a desire to perform A that results from the practical reasoning but an intention to perform A). The anti-causal theorist might ask how a desire for an end leads to a desire for something thought to be a means or to an intention to perform an action thought to be a means. If this connection is causal then it would seem that ~~with~~ the desire for D leads to the bodily movement in a way that is not entirely causal (because if we accept his basic claim the last stage in the proceedings is not causal), or the desire for D causes the bodily movement directly, thus by-passing the desire for the means or the intention. It does seem that, at least in some cases, the process of one desire leading to another (or to an intention) is a process occurring in temporal stages. (7) And it seems implausible for the causal theorist to adopt the second alternative - that the desire for the end D alone causes the bodily movement, the belief that performing A is a means to getting D, being not involved in the process but in some unexplained way accounting for the rationality of the action. Furthermore, if it is said that the desire for the end together with the belief about the means lead directly to the action, the anti-causal theorist can point out that a logical connection between the former two entities and the action can be extracted by constructing a practical syllogism.

So it seems that if a causal theory of action is to be rendered defensible it will have to meet directly the claim that logical matching rules out causation. It cannot be done by restricting the theory to desires whose object does not mention the action performed nor by emphasising that the desire to perform an action causes a bodily movement rather than an action. As Davidson has pointed out the object of a desire to perform a certain action is not the particular act-individual which one does perform when one acts on the desire - for that action will have a host of properties not mentioned in the specification of the desire. If I want to go for a walk, and go for a walk, then if I say "This is the walk I wanted to go on", this can be viewed as an assertion that the walk fully meets the specifications expressed by my desire. But what could it mean to say that every aspect of the walk "met" (as distinct from "was compatible with") those specifications, that no aspect of the walk could have been otherwise without the walk's failing fully to satisfy my desire? (8) This can be expressed by saying that if someone desires D (where D is a state of affairs) the description D does not refer to a



particular state of affairs or event or fail to do so - the question of uniqueness of reference does not arise. It might refer, in some sense, to a class of states of affairs or events, but to say this is to admit the point at issue. One can, of course, want objects but as Kenny, among others, has pointed out one must always be prepared to say what counts as getting the object, and possibly what one wants it for, so this does<sup>not</sup> constitute an objection to our thesis. (9)

Clearly, the connection between "A desires D" (an event or state of affairs) cannot be one of implication. "A desires D" does not entail that D comes about. Nor does it entail that D is not at present the case, although it does seem to be true that if someone desires D (straightforwardly interpreted) he cannot rationally believe that D is already the case. (10) So the anti-causal theorist's claim amounts to the thesis that desire and action cannot be causally related because there must be descriptions which match, but not because of any entailments from statements affirming the fact of the desire and statements affirming the existence or non-existence of the state of affairs desired, nor because the existence or non-existence of the state of affairs is referentially presupposed by the statement that someone has the desire. So we are left with the blank assertion that the necessity for descriptions which link the desire with the action rules out the possibility of a causal connection. Where can we go from here? We can argue that the claim rests on a misinterpretation of the well-known Humean requirement that a cause and its effect must be logically distinct (as is done persuasively by Goldberg (11)) and we can point out the existence of descriptions of the desire and the action which do not even superficially match in any verbal or logical sense (for example, the desire someone felt at time<sub>1</sub> and his eating a biscuit). Or we can demand some reason for insisting on the thesis that the necessity for descriptive matching rules out a causal connection.

Finally we can look for positive arguments against the thesis being true. We have spoken of the necessity of there being a descriptive match between the reason and the action. Necessary for what? For the reason to explain the action as an action done for that reason. So a certain requirement must be met

if there is to be a certain pattern of explanation. But that this requirement is met is supposed to rule out what is considered to be a pattern of explanation along other lines, namely causal explanation. But this cannot well be as matching simply does not rule out causation. (There is some degree of descriptive matching in "I applied a match to the gunpowder" and "The gunpowder exploded"). So the descriptions must match in a certain way, if causation is to be ruled out, and this type of matching is that required for a reason to explain an action as an action for that reason. There will often be descriptions of the cause which link it with the effect. The difference here, the anti-causal theorist will claim, is that there must be such matching descriptions, if the explanation is to be one of explaining an action by giving a reason for it. But why should this modal property rule out causation? It is not that the descriptions must match simpliciter, but that the descriptions must match if the explanation is to be of a certain sort. There seems no reason to suppose that the fact that a requirement must be met if an explanation is to be of a certain sort, ascribes some property to the terms of such an explanation which rules out their being corrected in some other way (if causation is another way). If the fact that the descriptions offered of the explicans and the explicandum can match without this being a case of explanation of an action in terms of reasons for acting, does not rule out causation then it does not seem that the requirement that they must match if the explanation is to be that of giving a reason for acting, can do so. (12)

I have earlier given an argument to the effect that if action and reason for action must be, in some sense, logically related, this logical relation is not one of entailment. It might seem that I connect the possibility of a causal relation between desires and the actions they explain as reasons with the intensionality of desire. And the intensionality of desire-statements does indeed help to establish the possibility of a causal connection in such cases. But it should be noted that events whose descriptions are both extensional and match closely can stand in causal relations. For example, the kind of mechanisms known in science as "fee<sup>d</sup>-back" mechanisms are of this kind. A thermostat is an arrangement whereby the event of the temperature in a region rising above a given level sets in operation a mechanism which causes it to fall below that level. And I suppose

one could have an arrangement whereby the event of the temperature in the region rising to a certain level sets in operation a mechanism which caused it to be maintained at that level. Because of the analogy (the strength of which I shall not attempt to assess) between such processes and perception followed by action such mechanisms have been favoured as models for the mechanistic and quasi-mechanistic explanation of human behaviour. (13)

Now I have so far spoken in a way which does not make it clear how I think that desires are causally involved in actions. It will not do to maintain that desires cause actions as this leaves the concept of action unanalysed, and we are trying to say what it is to act on a desire. As far as I can see such an account would block any analysis of action, unless one says that actions are those bodily movements which in the right circumstances are caused by desires. The thesis must be that desires cause bodily movements, and such processes constitute actions (the action cannot be identical with the bodily movement as this would have the absurd consequence that events in the spinal cord, not even in the brain, could be causes of the action (14) unless one is prepared to deny that if A causes B and B is identical with C, then A causes C). For the time being I want to ignore mental actions, acts of omission and acts<sup>of</sup> refraining where there is no bodily movement, but there is an action, or at least an act. But now we must face the argument mentioned above: if the action of raising one's arm is analysed as the desire to raise one's arm causing one's arm to go up, then the "desire to raise one's arm" will have to be analysed as "the desire that the desire to raise one's arm should cause one's arm to go up" and we seem to be involved in self-reference. I do not know of any explicit statement of this argument but it can be seen as a development of one given by White in his introduction to "The Philosophy of Action". (15) Speaking of the attempt to explicate an intentional bodily movement as a bodily movement preceded by an intention, he claims that preceding intentions cannot convert bodily movements into actions as what they can properly be said to precede are actions and not bodily movements. To the counter that it is because the bodily movement is preceded by an intention that we can or must change to the terminology of action, he might reply that the idea of an action is already contained in the intention to perform it and still awaits analysis.

In some cases the desire in question might be that one's arm should go up and not that one should raise one's arm, but true or false, this claim does not take care of cases where it is a desire that one should raise one's arm that is efficacious. I think that the best way to handle this difficulty is to take our lead from Davidson's point that the object of a desire to perform a certain act is not the particular act-individual that one performs because of it. We can say that the desire to perform a certain action, when this means more than that a particular bodily movement should occur, is not the desire that the desire itself should cause the bodily movement but that some desire should cause the movement (in the case of what Frankfurt calls "second order volitions", of which more later, it might be that some other desire should cause the movement), (16) that is the feature of the act-individual which one performs because of the desire, namely that the bodily movement involved in the action is caused by the desire is not mentioned as part of the object of the desire.

### Chapter 3

#### The Various Ways of Explaining Actions

The discussion so far has centred upon the question as to whether acting on a desire can possibly be analysed as the desire's causing a bodily movement. But so far we have said nothing about whether there can be non-causal explanations of action. There are three questions:- (a) Is Davidson's causal analysis of the "because" in "He did A because he desired D" the only possible analysis? (b) Can there be explanations of the form "He did A for the reason R" where the reason does not involve desire, and if so, is the explanation still causal? (c) Are there any other kinds of explanation of action other than those mentioned, and if so, are they still causal? I shall deal with (c) first. Thalberg in his book "Enigmas of Agency" (17) defends the view that "there is a distinction between explaining an event by fitting it into a general cause-effect pattern and another manner of making the event intelligible which we dub 'essential explanation'. Roughly, to propose an essential account of some incident is to delineate those qualities or aspects of the incident which figure in our criteria for saying what kind of occurrence it is. You could later go on to give an essential explanation of the kind of occurrence itself; this would be philosophical analysis...The difference in levels should be obvious. First you might say what makes your wife's activities in the kitchen this afternoon an instance of following a recipe. Then you might elucidate the notion of following a recipe, showing how it is connected with the more general notions of obeying rules, skills and knowledge." (18)

He gives the following illustration - "The situation is that during a recent storm in the vicinity of Chicago many buildings were blown apart, while immediately adjacent buildings remained unharmed. Why? An essential explanation of the incident could run: the storm was a tornado and tornadoes consist of violent winds of small extent that touch the ground only in certain places along a narrow path...the account you received does not list preceeding events in standing conditions which brought about the violent winds, nor does it cite laws of nature to explain how these winds demolished buildings. Seeing the incident as a type of wind-storm that blows down things along a narrow path is not at all the same as discovering that it exemplifies

a causal regularity." (19)

Later on, Thalberg considers the case of "a noble, savage uncontaminated by society. He comes upon a beach, notices another noble savage drowning and pulls him ashore." Thalberg continues, "An essential explanation of the occurrence...would settle two questions: 'why does the episode count as an action by the rustic rescuer, instead of being either the action of some other participant, such as the drowning man, or else not an action at all but possibly something that happened to one or more of the participants?' and 'What marks the incident as the specific act of rescuing?'" At this juncture, reasons and particularly desires move to the centre of the stage. We could propose either of two mutually compatible essential accounts which provide at least logically necessary conditions: (3a<sub>1</sub>) The episode is an action because it is under the rescuer's control; more vaguely it conforms to his desires at the time. (3a<sub>2</sub>) The episode is an action by our unpolished protagonist because it is caused by his desires and beliefs." (20) Further on, Thalberg says "that it is by reference to his desires that the native himself understands his behaviour. Moreover, his view of what he did, as defined by his desire, has a special primacy for other people who report his action. In other words, the native's performance fulfils a purpose or intention that he could acknowledge as his own. He would describe what he was up to in terms of his purpose or intention, if he had occasion to be candid with us. Furthermore, it is this description of his behaviour, the one that he is disposed to give, when he is open and truthful, which makes what he did the type of act it is, rather than some different type." (21)

My comment on the foregoing is not necessarily in disagreement with Thalberg, but merely to point out that, on this showing, an essential explanation of an action seems to involve the agent's reason for an action and hence causality if the relation between reason and action involves causality. (Of course, as Thalberg rightly points out, to claim a causal relation is not to cite the causal law involved). And as far as I can see, any essential explanation of an action (and not the bodily movement involved) must mention the agent's reason if it is to be really an explanation of the action. And thus to offer essential explanations of actions as instances of non-causal explanations of actions is only plausible if there are reasons to believe that to explain

by giving the reason is not to offer a causal explanation. The reason why I do not think that I am necessarily in disagreement with Thalberg may be brought out by the following quotation: "Why does the essential explanation you derive from the agent's avowal fall short of a causal account? No doubt he too is ignorant of the causal uniformities which this case exhibits. The descriptions under which he reports his act and its cause probably do not appear in any accepted or plausible law-like generalization. His action and the event he is reporting when he announces his reasons may have to be reported in different terms, perhaps from another system of discourse, before the result and these antecedents can figure in a causal explanation of what he did." (22) Thalberg is thinking of a causal explanation as being the presentation of the relevant law-like generalizations whereas I am thinking of it as often involving only the citing of the cause, or even a cause. Causal explanation of a rough and ready sort also enters into Thalberg's essential explanation of the peculiar pattern of destruction in the Chicago storm. To accept the explanation, one has to assume that winds of a certain type will cause destruction of a certain type (a generalization).

Thalberg also points out that essential explanations often mention the existence of conventions, rules and institutions by virtue of which actions are what they are. For example, a life-guard's action in saving a drowning swimmer counts as "doing his duty" in virtue of the existence of the institution of employing life-guards. Or an explorer's action of entering a cemetery might count as "desecrating the cemetery" because of customs existing in the region. Not all the examples Thalberg would regard as giving essential explanations of actions seem to be explanations of actions. For example, to say what makes someone's action an instance of desecrating a cemetery is not to explain his desecrating a cemetery or to explain his action under some other description. It seems that an essential explanation of an action is an explanation of an action only if it makes reference to the intentions, desires etc. of the agent. Otherwise it is merely an explanation of why the action counts as an action of a certain type. Certain actions might count as desecrating cemeteries whatever the intention behind <sup>them</sup> it, and whatever the explanation of them i.e. in virtue of certain conventions there might be both intentional and unintentional acts of desecrating a cemetery. The essential explanation of the desecration of the cemetery might of course explain the behaviour of the inhabitants of the region towards the desecration, given

certain natural assumptions. But it remains true that to explain why an action is an action of a certain type is not necessarily to explain the action. (23)

Let us now turn to (b). Can there be explanations of action of the form, "Someone did something for reason R" where the reason is his reason, but is one that does not involve desire, and if so is the explanation causal? Natural suggestions include (i) actions performed out of fear (ii) actions performed in accordance with values of the agent, in obedience to a moral imperative or because he believes he ought to do something (iii) (to take an idea of Thomas Nagel's) actions for which there is a reason which is not a present desire of the agents - it might be a belief about a future desire or about someone else's desire. What seems to be common to these cases is that the attempt to introduce desires to explain how these reasons can motivate seems to involve a kind of redundancy. If we stipulate that the motivational efficacy of fear, one's conscience, values and reasons which are not present desires of the agent is mediated by desires to avoid what one fears, to act in accordance with one's values, to obey one's conscience or to satisfy the desires of another (or one's future self) we seem to fall foul of Wittgenstein's dictum; "a wheel that turns though nothing turns with it is not part of the mechanism". Why cannot fear, for example, motivate a person to take avoiding action directly? The example of fear is worth dwelling upon as a natural suggestion at this point is that fear itself involves desire. If someone fears that P, it might be said, this need only mean that he desires that not-P and thinks it likely or possible that P. This proposal has the merit that it neatly assimilates those fears which seem to be "calm passions", unaccompanied by bodily reactions or introspectible emotional feelings (24) and in so far as the emotional aspect of fear is not mentioned, it might be maintained that this is not the aspect of the fear that explains its capacity to motivate - in particular such bodily reactions as quakings and tremblings are not intentional actions.

There seems to be something a little odd about this analysis - namely, one wants to say <sup>that</sup> when one is afraid of something one's attention is directed towards that something and not its non-existence or non-occurrence. And even though, when one acts because of one's fear, one acts in a way designed to obviate its object, this does not mean that the idea of preventing that which is feared is



part of the fear itself - it may appear as a result of the fear or in the belief that, together with the fear, leads to action. Furthermore, if the suggested analysis is permissible, why not reverse the process and analyse desire in terms of fear? If someone desires P, he merely believes that P is not the case, and fears that it may continue not to be the case and (if we adopt one form of the distinction between a desire and an idle wish) believes that P is a real possibility.

This last point is an important clue to how we might deal with the general problem of the reasons for action which possibly do not involve desire. Fear might be capable of motivating action towards avoiding the object of the fear, just as desire is capable of motivating action towards bringing about the object of the desire. And if someone has a certain fear (say that P) a certain desire (that not-P) becomes ipso facto equally intelligible: and if someone has a certain desire (say that P) a certain fear (that not-P) becomes ipso facto equally intelligible. (Actually I am ignoring a significant ~~as~~<sup>so</sup> asymmetry here: someone may well desire P and it may be obvious to him that it is within his power to bring it about that P, <sup>so</sup> as there would be something unreasonable in his fearing that P should not come about. For this reason, if one is bent on reducing the one to the other, it is rather more plausible to analyse fear in terms of desire than desire in terms of fear. But if we know nothing of the man's beliefs about the possibility of P, then given that we accept that he desires P, a fear that P will not come about would seem reasonable). In so far as a fear that P motivates an action, it will always be possible to construct a parallel case in which a desire that not-P motivates a similar action and the action will seem neither more nor less intelligible. Thus if a certain fear that P is not reducible to a desire that not-P, and if it can motivate an action without the mediation of some other desire we need not think of this as a case radically distinct from desire-motivated action. To say that someone did A because he feared that P is to say something that still justifies the Davidsonian (25) demand for an explication of the word 'because' and in so far as the action is intelligible we must assume that the agent believed that A was a means to bringing about that not-P (or a way of averting the feared consequences of P. It is worth noting how fears lead to further fears - I come to believe that Q will bring about that P which I fear, so I therefore come to fear Q. This should be compared with how desires lead to further

desires by means-end reasoning. There ought to be a certain form of practical syllogism appropriate to action out of fear). These requirements are also present when someone does A because he desires that not-P. It is hard to see how a fear that P can motivate action in a manner very different from the way that a desire that not-P motivates a similar action.

What then of the motivational efficacy of moral imperatives? Of moral values and beliefs about what one ought to do? Can they be construed as desires or something very like them? It is not clear how a moral imperative or the voice of conscience if this is interpreted literally as a kind of command (perhaps an inner command) can at the same time be viewed as a kind of desire or something very like a desire. But if conscience is viewed in this way, it is not idle to introduce a desire to obey it. If X gives Y a command, it is not sufficient for the command to be obeyed that a bodily movement occur - an action must occur. And further, a distinction could be drawn between an action which merely accords with the command but is not performed because of it, and one that is performed in obedience to it. Thus if conscience is construed as an inner voice, issuing a command (I am not concerned with the merits of the theory), then to say that someone obeyed the dictates of his conscience is to say that he performed an action which he <sup>intended</sup> viewed to accord with the command - and here it is not otiose to introduce a desire to obey the command, in the way that it seemed otiose to introduce a desire to avoid what one fears to mediate between one's fear and one's action. Might it be sufficient to introduce an intention to obey a command? This would not make much sense of the action of obeying one's conscience, unless it were a long-standing intention to obey the dictates of one's conscience, which it would be difficult to represent as not stemming from a desire. It might very naturally be said that the voice of conscience itself provides one's reason for obeying it if one does but this is to abandon the model of obedience to one's conscience as obedience to a command in any simple form.

If action in accordance with one's conscience, a moral imperative or a belief that one ought to do something is construed as a reason (conscience, belief, imperative) leading directly to action without the need for a further reason for acting in accordance with the reason, then it seems to me that there is still reason to think that we are dealing with a case of motivation that is closely

analogous to motivation by desire. Whether moral values can be analysed in terms of desire is a question I shall not pursue here (26) but when someone acts in accordance with his moral values the episode resembles action in accordance with desire in the same way as did action out of fear. To say that someone did A because he believed he ought to do A is to say something concerning which the Davidsonian demand for an explanation of the word 'because' is appropriate. <sup>And</sup> If someone believes that he ought to bring about P and performs A because of this, then if the action is to be intelligible we must assume that he believes that performing A is a means of bringing about that P.

It seems reasonable in the light of such considerations to endorse Kenny's suggestion that we should recognise a genus of pro-attitudes (in all cases the agent, as he puts it, volits that P) and in so far as these can motivate action (perhaps not all of them can - possibly mere contentment with the present situation cannot, and only such things as desire to prolong it can do so) to assume that they do so in the same way. We can then take the case of desire as typical. If someone does A because he believes he ought to do A, this is like doing A because he wants to do A. With fear it was possible to say that actions done out of the desire that P and actions done out of the fear that not-P were, with one restriction, equally intelligible. In the present case we cannot say this as someone's doing A because he wants to might only be intelligible if we take this as meaning "because he believes he ought to". Desire that P and fear that not-P seem to allow the same possible values for 'P', whereas moral values seem more like a sub-class of desires, or of the more general genus of pro-attitudes.

It might seem paradoxical to bracket belief that something is good or ought to be brought about with desires rather than other beliefs but if we remember Miss Anscombe's distinction between the direction of fit appropriate to desire and that appropriate to belief it seems that value-beliefs have the same direction of fit as desire. Whereas we judge beliefs according to whether they fit the world, we judge the world, that is, states of affairs, events, actions, according to whether it fits our desires and values and take action accordingly. This does not mean that desires and values are beyond criticism, but that in addition to ~~any~~ susceptibility to assessment they may themselves possess, they also have another direction of fit in that the actual state of affairs can be judged

according to whether it matches them. Let us imagine someone whose philosophical <sup>convictions</sup> ~~valuations~~ lead him to assimilate all desires to beliefs that something is good - he would still have to make the direction of fit distinction within this more general category of beliefs if his theory were to make any sense of action.

I have not yet dealt with Nagel's (29) reasons for action which are not present desires of the agent. His argument runs as follows:- "The assumption that a motivating desire underlies every intentional act depends, I believe, on a confusion between two sorts of desires, motivated and unmotivated...many desires, like many beliefs, are arrived at by decision and after deliberation... The desires which simply come to us are unmotivated though they can be explained. Hunger is produced by lack of food, but is not motivated thereby. A desire to shop for groceries, after discovering nothing appetising in the refrigerator, is, on the other hand, motivated by hunger. Rational or motivational explanation is just as much in order for that desire as for the action itself. The claim that a desire underlies every act is true only if desires are taken to include motivated as well as unmotivated desires, and it is true only in the sense that whatever may be the motivation for someone's intentional pursuit of a goal, it becomes in virtue of his pursuit inso facto appropriate to ascribe to him a desire for that goal. But if the desire is a motivated one, the explanation of it will be the same as the explanation of his pursuit, and it is by no means obvious that a desire must enter into this further explanation. Although it will no doubt be generally admitted that some desires are motivated, the issue is whether another desire always lies behind the motivated one, or whether sometimes the motivation of the initial desire involves no reference to another unmotivated desire. Therefore it may be admitted as trivial that, for example, considerations about my future welfare or about the interests of others cannot motivate me to act without a desire being present at the time of action...But nothing follows about the role of the desire as a condition contributing to the motivational efficacy of these considerations. It is a necessary condition of this efficacy to be sure but only a logically necessary condition. In fact, if the desire is itself motivated, it and the corresponding motivation will be possible for the same reasons. Thus it remains an open question whether an additional unmotivated desire must always be found among the conditions of motivation by any other factors whatsoever. If considerations of future happiness can motivate by themselves, then they can explain and

render intelligible the desire for future happiness which is ascribable to anyone whom they do ~~not~~ motivate." (30)

Later on he writes - "If the act is motivated by reasons stemming from certain external factors and the desire to perform it is motivated by those same reasons, the desire obviously cannot be among the conditions for the presence of those reasons...the temptation to postulate a desire at the root of every motivation is similar to the temptation to postulate a belief behind every <sup>in</sup>ference...this is true in the trivial sense that a desire or belief is always present when reasons motivate or convince - but not that the desire or belief explains the motivation or conclusion or provides a reason for it. If someone draws conclusions in accordance with a principle of logic such as modus ponens, it is appropriate to ascribe to him the belief that the principle is true; but the belief is explained by the same thing which explains his inferences in accordance with the principle. The belief that this principle is true is certainly not among the conditions for having reasons to draw conclusions in accordance with it. Rather it is the perception of those reasons which explains both the belief and the particular correlations drawn." (31) In order to argue that the view that prudential reasons (I shall deal only with these and not with altruistic motivation) stem from a present desire for the satisfaction of future desires and interests, is incoherent, Nagel argues against two assumptions on which this view depends. "First, the assumption that a desire or other relevant consideration can provide a reason only when it is present. Second, the assumption that any desire with a future object provides a reason for pursuing that object...Against the neutral view that a covering prudential desire is operative I contend first, that it does not take care of the actual cases (i.e. what we can explain in terms of prudence and the actual prudential reasons which we believe to obtain); second that the cases it does accommodate are not handled in the right way; as that their motivational nature is obscured by the theory..." (32)

The thesis that explains prudential conduct by saying that my future interests give me reasons to act because I have a present desire to further my future interests, Nagel writes, allows the following possibilities:-

- (a) "First, given that any desire with a future object provides a basis to do what will promote that object, it may happen that I now desire for the future something which I shall not and do not expect to desire then, and which I believe there will be no reason then to bring about. Consequently I may have reason now to prepare to do what I know I will have no reason to do when the time comes."
- (b) "Second, suppose that I expect to be assailed by a desire in the future... in the absence of my further relevant desire in the present, I may have no reason to prepare for what I know I shall have reason to do tomorrow."
- (c) "Third, expected future desires whose objects conflict with those of my present desires do not in themselves provide any present countervailing reasons at all...I may have reason to do what I know I will later have reason to try to avoid and will therefore have to be especially careful to lay traps and insurmountable future obstacles in the way of my future self." (33)

"Postulation of a prudential desire does not deal satisfactorily with the problems which I have imagined arise in the system without it. First of all its formulation presents serious problems. Presumably the prudential desire is supposed to yield a result based on the consideration of all other desires...past, present and future. It should obtain the conclusion by striking a balance between claims from different times. However it is simply one of the present desires and operates as such. So if one of its objects is the satisfaction of those present desires other than itself, they will enter the calculation of reasons twice; once in their own right and once as objects of the prudential desire. To avoid the result, the objects of the prudential desire could have to be restricted to future satisfaction. But this would not be satisfactory either, for a further balancing mechanism would then be necessary in order to settle conflicts between considerations derived from ordinary present desires and those derived from the future via prudential desire. Either this mechanism would be a further desire, in which case the same problems would arise all over again, or else it would be a structural feature of the system of reasons, in which case the project of accounting for prudence in terms of desires would have to be abandoned."

"Secondly, even if the problem could be surmounted, and there were a prudential desire, its presence would not alter the fact that the system through which it operates permits the derivation of reasons for action from any desire with a future object - not

only the prudential one...Therefore it remains the case, even if the prudential desire is present, that other desires with future objects can provide one with reasons to bring about what I know I shall have no reason to want when the time comes. This will occur if I have a desire for a future object to which I shall in the future be entirely indifferent (and about which the prudential desire is therefore neutral)."

"...if it (the prudential desire) has among its objects the satisfaction of all future desires, this could include the satisfaction of a desire for a still more future object. So if on Monday I expect that on Tuesday I shall want to eat a persimmon on Wednesday, although I also expect that on Wednesday I shall be indifferent to persimmons (as I am on Monday) - then on Monday I have a clear prudential reason to make sure I have a persimmon available on Wednesday, though I will not have any reason to want it then and I do not on Monday want to have it then." (34)

Finally Nagel writes;"...we must raise the question whether desires with future objects ever give rise to important reasons at all...I have already expressed doubt that desires are the most important sources of motivation. I now wish to extend this doubt, with particular emphasis, to the case of unmotivated desires with future objects...I am not talking about the motivated pursuit of future goals for independent reasons...Suppose that for no reason having to do with the future, I conceive now a desire to become a policeman on my thirty-fifth birthday. If I do not believe that the desire will persist, or that any circumstances then obtaining will provide me with reason for being a policeman, is it possible to maintain nevertheless that the desire itself gives me reason to do what will promote its realization? It would be extremely peculiar if anyone allowed himself to be moved to action by such a desire or regarded it as anything but a nervous symptom to be looked at with suspicion and got rid of as soon as possible." (35)

I find Nagel's negative thesis that it is unclear that we always need to introduce an unmotivated desire to explain the presence and motivational efficacy of a motivated desire more compelling than his thesis that the introduction of a present prudential desire to explain prudential motivation cannot do the work required of it. It is true that the thesis that any present desire with a future object provides a basis to do what will

promote that object gives one reason to prepare to do now what one knows one will have/<sup>no</sup>reason to do when the time comes. But one would not necessarily judge a man irrational merely because he had such a desire; one might only do so if he acted upon it - if it were not outweighed by countervailing present desires. And even if one did judge him irrational for having the desire, this would be to deny that he had a reason to promote the future object only in the sense that one did not oneself recognise the reason as a good reason, not to deny that he had a reason simpliciter. For suppose he acts on the desire - one does not want to say that he acted for no reason: that Nagel's procedure of asking whether a desire gives one reason to act (rather than whether it is a good reason) leads to this conclusion. If someone has a certain desire, then given certain beliefs, he ipso facto has a reason for certain actions, and one would invoke this sense of "reason" if he acted on the desire and one were asked for his reason for doing so.

I might say of someone that he has reason to perform a certain action. If asked what this reason is, I might mention desires and beliefs of the agent, and such an account could explain why he performed that action (if he did). Alternatively, I might mention factual considerations of which he is ignorant, needs of his of which he is unaware or desires which I think he ought to have. Such an account would also be a case of saying that someone has reason to perform a certain action. But it could not, of course, explain his performing that action (if he did). Similarly, if I expect to be assailed by a desire in the future, in the absence of any relevant present desire, I may have no reason to prepare for what I know I shall have reason to do tomorrow. But this is only counterintuitive if it is taken to imply not only that I have no reasons for the action (I could not give any reasons which are genuinely my own) but also that a second party could not say that there are reasons for me to perform the action. The sense in which it might be true that I have no reason to prepare for what I shall have reason to do now, in the absence of any relevant present desire, is the one which might have to be invoked to explain why I do not take any preparatory measures.

I am not arguing that Nagel is wrong to say that beliefs about one's future desires and interests can motivate; but that the denial of Nagel's thesis does not involve one in the counter-intuitive conclusions he thinks it does. If someone has a belief



about his future desires but does not act on it and believes he can do so, then if there are no other countervailing desires present, if we do not adopt Nagel's view, one can explain this by saying that he has no prudential desire or that if he has the prudential desire it did not interact with the belief to yield a present desire to perform some preparatory action or that the desire to perform the preparatory action, although it came into existence, simply did not lead to action. Nagel, on the other hand, would just have to say that he had a reason to perform the preparatory action but did not act on it. In the first chapter, I defended the view that it is possible to have a desire, believe it is possible to act on it and yet not do so in the absence of countervailing desires. Nagel will have to adopt a similar view as regards having reasons in his sense, and it looks as though the failure to act on them even when there is no explanation in terms of countervailing reasons or the absence of relevant beliefs, will for him be not merely a possibility but a regular occurrence (perhaps not as regards prudential reasons, but probably as regards altruistic reasons). It would be implausible to maintain that if someone claimed to believe that someone else desired something yet in the absence of countervailing reasons etc. did nothing about it, this should be treated as a criterion that he did not really have the belief. And such cases might be common).

Returning to the arguments given on pages 32-33 of this chapter, I would accord a similar treatment to Nagel's statement that on the view he is attacking "expected future desires whose objects conflict with those of my present desires do not/<sup>in</sup> themselves provide any present reasons at all". Lest it be thought that my introduction of a second party, someone who judges the agent's reasons as good or bad, begs the question in favour of some kind of subjectivism or relativism, that is does not allow that someone can have reason to do something unless he has certain desires, or someone else believes he ought to have certain desires etc. it should be pointed out that, if there are objective reasons in this sense these cannot motivate unless someone has them in the strong sense of having the relevant beliefs and desires. There may be reason for me to do X irrespective of what I or anyone else believes or desires, but such a reason could not explain my doing X irrespective of what I believe or desire. Rational action is not Reason acting through one but oneself acting in a way that can be justified by reference to reason.

Nagel's argument that the formulation of the prudential desire

presents difficulties seems to depend on the assumption that the prudential desire is supposed to arbitrate between one's desires and somehow arrive at a verdict concerning their respective claims. But this is hardly the purpose for which it is introduced. Given knowledge of a future desire, the prudential desire is postulated to explain how this knowledge can figure in motivation and presumably it does so in the same way as other beliefs figure in motivation - they are relevant to the way in which a certain desire can be satisfied (in this case the prudential desire, that is the desire that one's future desires should be satisfied). It is not intended to determine whether the prudential reasons will win out, or whether they ought to. Thus the prudential desire together with the belief about a particular future desire leads to a present desire to satisfy that particular future desire. As he himself points out, it is difficult to see how any desire could have the feature which Nagel thinks the proponent of a present prudential desire could have to ascribe it - of settling conflicts between desires. One could have a desire that some desires should always be overruled - but this would be a desire amongst others and it seems that it could only have privileged status in the de facto sense that it might always prove triumphant. It is not clear to me why Nagel should demand an account of how desires or reasons balance out at this juncture or what kind of answer he expects. He suggests that "the mechanism would be a structural feature of the system of reasons, in which case the project of accounting for prudence in terms of desires would have to be abandoned." (36) This is obscure to me unless it means that it is rational to take into account prudential reasons and ( if the remark is to have any bearing on the settling of conflicts between them) it is rational to give them a definite weight in any particular case. But then the account looks as if it is one of how someone must choose if he is to choose prudently or rationally and not how it comes about that he makes the choice he does: the latter is not a normative question and further more must deal with someone who does not make a prudent choice even though he has prudential reasons, whether or not these are present desires. But I cannot see why either the normative question or that of the mechanism by which an agent's conflicting reasons yield a conclusion ( whether rational or otherwise ) in action should be raised here. (37) Prudential considerations can motivate but need not do; all we are trying to explain here is how they motivate in those cases in which they do.

Nagel further claims (i) that if there were a prudential

desire the system through which it operates would permit the derivation of reasons for action from any desire with a future object - not merely the prudential one and (ii) that if the prudential desire has among its objects the satisfaction of all future desires, this could include the satisfaction of a desire for a still more future object, thus providing a prudential reason to make sure I get at this still later time what I do not want now, to have then and I do not expect to want them, can be met in the same way as before. A person with such desires does have reasons in a sense which is not counter-intuitive, the sense we would invoke if he acted on the desires and we correctly gave these desires as the reason why he acted, but this implies nothing about the reasonableness of his desires, or of his acting on them in a particular case or about whether we or anyone else endorses them.

Nagel's claim that unmotivated desires with future objects are relatively unimportant seems to rest on the view that most of them are unreasonable. One could of course accept this without denying that they are reasons of the agent in question, and if he acts on them his reasons for doing things. But is it true that present unmotivated desires with future objects are either unimportant or usually unreasonable? If someone wants to go to the South of France next summer, we can perhaps say that this is a motivated desire - what he really wants is a holiday in a warm climate next summer and has appropriate beliefs about France. But the desire to which we have traced back the first desire is still a desire with a future object - it looks like a perfectly good unmotivated desire with a future object. Such desires seem neither uncommon nor unreasonable. Perhaps Nagel will say that the man does not really desire to take a holiday in a warm climate next summer but believes he will want to then. This might be the case, but need not be. Or perhaps he will say that what he wants is to take a holiday now and is forced for various reasons to postpone the satisfaction of the desire. This is a little more plausible in that if someone specified a particular time in specifying his desires, it is usually in order to ask why he picks that particular time and perhaps this makes the desire a motivated desire. But it does not seem that if one traces the source of the desire back to an unmotivated one, one must end up with a desire with a present object, in this case a desire to take a holiday in a warm climate now. Perhaps physical desires such as hunger are of this type but with the desire to take a holiday in a warm climate next summer it could be that what one uncovers at its source is simply the desire to

take a holiday in a warm climate - without any temporal satisfaction.<sup>specific</sup>  
Or perhaps to take a holiday in the not too distant future.

In any case there is something odd in the idea of an opposition between desires with present objects and desires with future objects. A desire with a present object (if this is taken literally) would in many cases be rather like a desire with a past object - it could not motivate a rational man except perhaps to find out whether what he desired at  $t$  to be the case at  $t$  was the case at  $t$ . What Nagel intends by "a present desire", I suppose, is a desire for something as soon as possible. This is a limiting case of a desire for something in the near future. It seems arbitrary to oppose this type of desire (for something as soon as possible) to a category which lumps together desires for something in the not too distant future, desires for something some time or another, desires for something before a certain specified time, desires for something during a certain later stretch of time, desires for something at a more or less precisely defined time in the future, desires about what should happen if something else happens etc. No doubt some of the latter cases may, if unmotivated, seem bizarre. But this does not show that they all are and in particular that a prudential desire (i.e. a desire that desires which I might have in the future generally should be satisfied) is.

It does not seem therefore that Nagel has shown that the introduction of a present prudential desire cannot do the work required of it or does it in the wrong way. There are problems about the prudential desire, for example, concerning its individuation - does one simply have a standing desire to further one's future interests or does one sometimes have such a desire, and sometimes not, and is it the same desire each time? I shall not deal with such problems as my purpose is not to show that a prudential desire is the explanation of prudential motivation. And indeed it has not been shown that it is necessary to introduce such a desire to explain prudence. Nagel's alternative (39) is that it is "a formal feature of the system that there is a reason to do not only that for which there is a present reason, but also what will promote something future for which there will be a reason". If he is talking of how reasons can motivate he must be talking of reasons that someone believes to obtain. Thus, for example, he argues that a belief that one will desire something can itself motivate without the need to postulate a present desire

other than that which can be ascribed on the basis of the agent's intentional action (actually he locates the motivational content in a tenseless judgement about the future time in question - "that judgement possesses the motivational content possessed by the present tense version") (39)

If someone believes that at some future time he will desire X and believes that if he performs A he will secure X for himself at that future time, then we have a situation specified in a way that mentions no present desire. Suppose he performs A - Nagel points out that it is appropriate to ascribe to him a present desire to perform A, but this is a logical truth which licenses the ascription on the grounds that the action is intentional. The two beliefs might, it seems, be sufficient to cause the bodily movement. Unless one high-handedly makes the pointless stipulation that the two beliefs alone could only cause a bodily movement that is not part of an action, one is presented with two possibilities:-

- (a) The two beliefs might lead to action without the mediation of any desire
- (b) The two beliefs might lead to action with the mediation of a desire, but the presense of one of the beliefs alone accounts for the presense of the desire i.e. a belief that he will in future desire X leads on its own to a present desire for X.

At first sight it seems the question of whether either of these possibilities actually occurs is an empirical one. But there is a problem as what one would take as evidence of the occurrence or non-occurrence of a mediating desire as in (b) or of a prudential desire which would mean that we do not have a case of either (a) or (b), but of ordinary desire-based motivation. Does one rely on introspection, or hope that some evidence will turn up in the form of neurophysiological considerations? Let us take a closer look at (a), assuming for the moment that it is a well-defined alternative. Someone believes that he will in future desire X and believes that performing A will secure X for that future time. Without the mediation of any desire, he performs A with these two beliefs as his reason. Let us compare this with someone with a present desire for X at the future time and the belief that performing A will secure X for that later time and who performs A with this desire and this belief as his reason. In the former case the belief about the future desire seems to be functioning as if it were a present desire with the same object (the objects of the desire

including the same temporal specification in each case) I suspect that this boils down to the same point as Nagel's claim that the motivational content of a belief about a future is <sup>located</sup> treated in a tenseless judgement about the future time in question.

At this point we can apply the same considerations to motivation by beliefs about one's future desires as we did to motivation by fear, moral values etc. In any case of apparent motivation, namely by a belief about one's future desire, one can construct a parallel case in which this occurs by interaction of the belief with a prudential desire which is neither more nor less intelligible than the original case (this last clause shows why I laboured to refute Nagel's attack on the explanation of prudence in terms of a prudential desire. If he were right, then the parallel case we construct would be less intelligible than the original because of the alleged oddity of the prudential desire and of present desires with future objects). And again there is the relevance of the Davidsonian demand for an account of the "because" and the need for a matching belief about the action performed because of the belief that one will have a certain future desire.

I have deliberately left it an open question whether there are three distinguishable possible types of prudential motivation - via a prudential desire, and the types labelled (a) and (b) above. Let us dub (a) and (b) examples of bifunctionality, because a belief seems to be behaving not only as a belief but also as a desire, which then leads to action (in the same way as desire). Such beliefs are exhibiting two directions of fit. If I believe I will want X at a later time, my belief is wrong if I do not want X at this later time. On the other hand if such beliefs can motivate then one is adjusting the world to fit them and judging the world according to whether it fits them. Of course, the first direction of fit we compare the world with the object of the desire embedded in the belief.

Are prudential motivation where the motivation involves a present prudential desire and the two kinds of bifunctionality distinguishable? What a person is inclined to say about his beliefs and desires might suggest an answer in some cases. If someone reports a strong desire that any future desires he has should be satisfied, one might accept this and opt for an explanation of his prudential behaviour in terms of a prudential desire. Furthermore if someone says he believes he will desire X and does nothing about

it, and later says he believes he will desire X and desires that this desire be satisfied, and this time does something about it, we might say that belief alone was not sufficient to motivate him, but that a desire was required too. But why should these possibilities show anything about all cases of prudential behaviour? I cannot think of any clear-cut criteria for distinguishing the three possibilities, different though they seem to be on paper. As I mentioned before, neurophysiology might provide an answer, but is it certain that neurophysiology will one day enable us to read off from someone's brain state whether or not he has a certain desire and whether or not this desire or even this same desire is activating his present behaviour? Not even all versions of the mind-brain identity thesis maintain this. (40) Perhaps bifunctionality of type (a) has the merit of economy - it does not multiply entities beyond necessity. If it is not a clear necessity to introduce present desires to explain prudential motivation then why do so?

Finally we must deal with question (a) at the beginning of this chapter. (p.21) Is Davidson's causal account of the "because" in "He did A because of his desire for D" the only possible account? Suppose someone desires D and believes that performing A is a means to getting D. Suppose a bodily movement M of the type which could be involved in A occurs and suppose that this movement is not caused by the desire. Now a bodily movement is surely the sort of thing that can be caused. So the possibility arises that M has a cause other than the desire. Now some types of cause other than the desire would defeat utterly the view that the man had performed an action. (Suppose, for example, the movement were brought about by an outside manipulation, or were caused by some other irrelevant desire, say for E). This is not to say that they all would - in particular, some events in the nervous system would hardly have this effect. But it looks as though if one says that the desire from which an action proceeds is not causally related to the bodily movement, one can then do little more than list causes of the bodily movement which would tell against its being part of an action, and those which would not. One could indeed give a rationale for the groupings; but the trouble is that whatever relation did hold between the desire for D and M (say, that which can be exhibited by constructing a practical syllogism) this could never guarantee that one had an action, as one could always construct a case in which the relation held, but where M had causal antecedents which would make it wildly implausible to say that M was performed because

of the desire for D. A causal relation of some kind between the desire for D and M seems the only one which has any hope of being strong enough to rule out other causes of M which would be incompatible with M's being performed because of the desire for D. Obviously there will be great complications when we come to consider cases where both the appropriate desire and something we would normally regard as falsifying the attribution of agency both make a causal contribution. But maintaining a causal relation between the desire and bodily movement seems necessary as a first step; ideally in a case of agency on the basis of a desire it seems the desire and the appropriate belief should be causally sufficient for the bodily movement to occur, given that certain parts of the nervous system and musculature are functioning properly. Otherwise if something else was necessary in a given case it might undermine the attribution of agency. (41)

But should we demand a relation between the desire and the bodily movement which rules out the bodily movement's being caused by factors incompatible with agency (or at least to rule out such factors' being the sole cause of the movement)? If we do not, we just have to say: certain causes of bodily movements preclude the bodily movement's being involved in action: if the bodily movement has no such causes, then provided it is appropriate to a desire (or desire and belief) in the practical syllogistic way, it is involved in an action performed because of the desire. But this is no good. Someone can have two reasons for performing a certain action. Surely we recognise the possibility that he might perform the action for only one of the reasons. But what can the difference be on the above account from his performing the action for the other reason, or for both reasons? In asking for a relation between the desire and the bodily movement which rules out the possibility that the bodily movement has causes incompatible with agency, we are also asking for a way of characterising what it is for an action to be performed out of one desire rather than another, when it is appropriate to both. Or at least, it looks as though if one adopts the defeasibility view of actions, one has no way of making this distinction.

The defeasibility view has something to be said for it: it does not preclude us from giving some explanation of why certain causes of a bodily movement would be incompatible with that bodily movement being performed for a certain reason. But when formulated it seems totally inadequate to distinguish an action performed



for one reason from an action performed for another when the action has the required appropriateness to both. We obviously need a stronger relationship between the desire and the bodily movement. We can state the relationship of appropriateness shown by constructing a practical syllogism without even supposing the bodily movement to occur. And we can also state this relationship of appropriateness between a certain state of affairs and a man's desire when that state of affairs comes about in a way totally independent of his actions. If the object of a man's desire comes about in such a way, we do not say that the man has brought it about; but on the view under attack if the object of a man's desire comes about independently of his desire (it merely satisfies it) then we might well have to say that he has brought it about because of his desire.

What we need is a relationship that can hold between two distinct entities, the desire for D and M and the truth of the statement that this relation holds has still to be determined even when it has been established (i) that the desire for D exists and M occurs and (ii) that the relation of appropriateness of M to the desire for D holds. Of course, the relation of mere temporal succession has this property. But it would not serve to tell us whether an action was performed because of a desire when it had the relation of appropriateness to the desire. Is a causal relation the only relation that can do the trick? Suppose a relation R were found, other than a causal one, which linked the bodily movement with the desire more intimately in the case where the bodily movement were performed because of the desire, than in the case where it were not, even though it had the relationship of appropriateness to both. Perhaps in the latter case, the relation did not hold at all. Then it seems that in a case where this new relation holds between a desire and a bodily movement, the bodily movement could fail to qualify as one involved in an action performed because of the desire of one of the defeasibility conditions mentioned earlier. (42) So in order for a bodily movement to be performed because of a desire, it would have to be related to the desire in this new way and not fall foul of the defeasibility conditions. The procedure of distinguishing bodily movements that were mere bodily movements (even though they accorded with one's desires) from those involved in actions would be radically different from the procedure of distinguishing bodily movements that are performed from one desire from those performed because of another (even though they accord with both). And this, I submit, is implausible; surely one

wants a unitary account of the relation between a bodily movement and a desire when the bodily movement is performed because of that desire. In both cases one wants to base this distinction on the connection of the bodily movement with the desire, not just on its lack of a connection with other factors with militate against regarding the bodily movement as an action as is required by the theory under attack when concerned with the first distinction.

Suppose someone wants to perform a bodily movement M. M occurs and we have reason to suppose (i) that none of the defeasibility conditions hold (ii) that R does not hold. Thus the bodily movement passes the test which is supposed to distinguish it from a mere bodily movement uninvolved in action because it has the wrong causal antecedents, but fails the test of linkage with the desire. What are we to say here? Suppose he had two reasons for performing M, say desires for D and E (together with the relevant beliefs). But R holds between M and neither the desire for D nor the desire for E. It looks as though we could not regard M as involved in an action - for we could give no reason for it, certainly not the desire for D or for E. Now, if this does show that we could not regard M as involved in an action, it is difficult to see how we can consistently avoid the same conclusion when there is only one candidate for the reason for which the action was performed. So it is not sufficient that none of the defeasibility conditions should be satisfied if M is to be involved in an action even in such a case. R must still hold between a man's desire for D and M when he believes that performing M is a way of getting D, and M occurs, even when none of the defeasibility conditions are satisfied - if he is to have performed M. But it is difficult to see any rationale for this on the view under attack. Either we introduce an unacceptable bifurcation<sup>urca</sup> in the ascription of action, and demand that an extra requirement be satisfied when we are dealing with actions for which the agent might have had more than one reason, or we demand that the requirement be satisfied when there is only one possible candidate for the reason, and here it is obscure for what reason, other than consistency, this should be done. Surely, it is better to take as R as a relation which (contra hypothesis<sup>em</sup>) excludes a causal relation to those factors which, if they were causally linked to M, would defeat the ascription of agency. (43)

## Chapter 4

### Actions, Reactions and Bodily Movements

In this chapter I want to consider certain difficulties which arise when one attempts to demarcate the concept of action precisely, in particular to distinguish from mere bodily movements such as might occur in emotional reactions. In the first chapter we saw that the evidence that a man has a certain desire includes considerations about his actions, but might also include bodily reactions characteristic of the desire. As the problem of human action concerns action for a reason even where the reason is only with a certain artificiality regarded as a desire (see the previous chapter), I will take fear as an example for investigation. Fear clearly leads both to action and to bodily reactions. As we have seen, although it is probably not possible to regard it simply as a desire, in so far as it motivates there is no reason to think that it motivates in a way significantly different from desire.

Suppose a man is afraid of something - he begins to tremble and runs away. We are inclined to describe his running away as an action, but we would not describe his trembling as an action. Why is this? We could say that the man ran away in order to avoid whatever it was he feared but we would not say he trembled in order to avoid anything. Now consider an animal that is afraid - because of its fear, it freezes. When would we say that it was paralysed by fear using the idiom of inaction and emotional reaction and when would we say it was feigning death, or remaining motionless in order to escape detection? I do not think that the answer would always be obvious. Nor would it always be obvious if the animal were a human being. If a man terrified at the sight of his enemy stands stock still (his enemy has not seen him) he is more likely to say, "I was paralysed by fear", if he wanted to do something else. But here it would be intelligible to raise the question whether he was paralysed by fear - a mere passive victim of his fear - or whether he acted in his fear but could not admit it even to himself. And if he had no countervailing desires to do something else, nor any objection to thinking of himself as motivated by fear, it is not obvious in which cases we would say that he was paralysed by fear and in which cases we would say that his remaining still was motivated by fear.

Miss Anscombe (44) has defended the view that in some cases

there is not a sharp distinction between a mental cause and a reason (I should prefer to make this distinction in a way that does not imply that a reason for someone's action is not a species of cause). For instance, someone hangs his hat on a peg because his host says, "Hang up your hat on that peg". Miss Anscombe writes that it would not be correct "to say that this is a reason and not a mental cause because of the understanding of the words that went into accepting the suggestion. Here one would be attempting a contrast between this case and, say, turning round at hearing someone say Boo! But this case would not in fact be decisively on one side or the other; forced to choose between taking the noise as a reason and as a cause, one would probably decide by how sudden one's reaction was. Further, there is no question of understanding a sentence in the following case: 'Why did you waggle your two fore-fingers by your temples?' - 'Because he was doing it', but this is not particularly different from hanging one's hat up because one's host said, 'Hang your hat up'. Roughly speaking - if one were forced to go on with the distinction - the more the action is a mere response, the more inclined one would be to the word 'cause'; while the more it is described as a response to something as having a significance that is dwelt on by the agent in his account, or as a response surrounded with thoughts and questions, the more inclined one would be to use the word 'reason'. But in very many cases the distinction would have no point."

Miss Anscombe is concerned with the limits of application of a distinction which is not quite the same as ours. She allows that mental causes are possible for actions as well as for emotional reactions. And furthermore, the suddenness of the response, which has some value in the example she is discussing, does not yield a clear answer in the case that interests <sup>us</sup> ~~me~~ - one could suddenly or slowly freeze in response to danger, but in neither case is the question whether we are dealing with an action or a reaction settled. We are assuming in our case that the man believes that by remaining still he will avoid alerting his enemy to his presence. Now if his remaining still is motivated by his fear, this belief must be involved in the genesis of his action, whereas if his remaining still is a case of "paralysis by fear", the belief seems inessential to the process, if not actually excluded from it. This tells us something of what we are affirming when we opt for one description rather than another. It does not imply that the agent always knows whether he acted on his fear rather than because paralysed by it, and thus always knows whether his belief was involved in the genesis

of his inaction. The existence of puzzle-cases need not mean that we cannot give conditions for the allocation of behaviour to the sphere of action or to that of emotional reaction. But it might often be that it is unclear as to whether or not these conditions obtain. With animals this unclarity can be glaring because of the difficulty of knowing when to ascribe beliefs to them, let alone when to regard these beliefs as involved in the genesis of their behaviour. It is noteworthy that teleological idioms come naturally in the description of much of the behaviour of animals which seem to be merely emotional or instinctive reactions. I suppose this tendency comes from the fact that we can often give a reason for the animal's behaviour in terms of its function in preserving the individual of the species. A biologist would no doubt frown on such a description in terms of function, and could offer one along Darwinian lines. And perhaps the temptation should in the last resort be resisted when we feel unable to ascribe the relevant beliefs to the animals.

If the causal theory of action is correct, a desire or fear might cause a bodily movement, and this sequence constitute an action. But a desire or fear could also cause a bodily movement and this constitute a reaction (I would suggest that "I trembled because I was afraid" has to be analysed causally to distinguish it from "I trembled and I was afraid". The trembling might have been because of a drug. Davidson's well-known argument seems to have a wide application (45) ). Now can we say that when a desire or fear leads to action this is by interacting with a belief whereas when it leads to a bodily emotional reaction this is not so? Consider the following definition of a basic-act type given by Goldman (46): (I am not here concerned with the adequacy of the notion of a basic act or Goldman's method of individuating actions):-

"Property A is a basic act-type for S at t if and only if

- (a) If S were in standard conditions with respect to A at t, then if S wanted to exemplify A at t, S exemplifying A at t would result from the want, and
- (b) The fact expressed by (a) does not depend on S's level-generational knowledge nor on S's cause and effect knowledge, except possibly the knowledge that his exemplifying A would be caused by his want".

The relevance of this lies in the suggestion that some acts are produced by desires (or fears) without the need for a matching belief. Suppose that every desire or fear required a matching belief

if it were to lead to action. We cannot then assume that the desire plus belief always leads to action via the formation of another desire (eg. the desire for D and the belief that performing A would result in getting D, leads to action via the formation of a desire to perform A) as this would be regressive. We must assume that the desire plus belief leads directly to action in some cases. Alternatively, we might take seriously the view that desires or fears might lead directly to action without the need for a matching belief. This seems appropriate to animals. If an animal desires to eat and sees another animal, it might then desire to eat that. The genesis of this new desire from the first might be interpretable in terms of some kind of interaction with a belief about the animal it sees being of a certain kind. But when the animal goes in to attack it seems artificial to introduce further beliefs about its acting thus leading to its being able to eat later. And if I desire to raise my arm is it really necessary to invoke a belief that I can raise my arm; to characterise the subsequent motion of my arm as part of an action? And even if one does not want to invoke such a belief (to say, for example, that a person must normally know what his repertoire of basic actions is ) this is not to say that the belief is involved in any particular action.

This complicates the problem of how we are to distinguish between action and emotional reaction. A desire for food causes someone to salivate, or his fear of an enemy causes him to tremble. How do we know that we are not dealing with actions unmediated by belief rather than emotional reactions? If we say that in the case of action one can always construct a parallel case in which the absent belief is present we face the following difficulty - imagine a man who is able to salivate at will and who believes that by salivating he can somehow increase the chances of his getting food, or imagine a man who can tremble at will and who believes that, by trembling, he can somehow avert the danger. How are we to distinguish such cases from reactions? (47)

It is a different sense in which a bodily reaction is appropriate to a desire from that in which an action is appropriate to a desire. An action can be appropriate to a desire because it is an action-type specified by the object of the desire, or if it is believed to be a means to satisfying the desire. If someone were able to salivate at will, and if we knew that he was hungry we would, if we regarded this as an action motivated by hunger, have to ascribe to him a belief that salivating would somehow lead to his getting food.

In order to regard it as a bodily reaction or change characteristic of the desire, this would not be necessary. So this is not a case of deciding between a reaction and an action unmediated by belief.

In order to produce a doubtful case, we would need one where someone desired to perform a bodily movement M and the right kind of bodily movement M occurred but where it was unclear whether M was involved in an action or was merely a characteristic bodily symptom of the desire. But it seems that in such cases we would always opt for the description in terms of action (48) (possibly an instinctive or reflex action). The upshot of this then is that if a desire causes a bodily movement this is an action if the bodily movement is of the type specified by the object of the desire, or if the bodily movement is believed by the agent to be such as to bring about the object of the desire and this belief plays a causal role in the production of the bodily movement. Although a desire might lead to action without the need for a mediating belief this type of case is already distinguished from reaction by the fact that the bodily movement is that specified by the object of the desire. Otherwise it is not and may be a bodily change characteristic of the desire or an emotional reaction. For fear to motivate it seems that one must always have the relevant belief or rather there must always be a gap between the bodily movement and the avoidance of that which is feared.

Our analysis does not imply that running away could never be merely an emotional reaction, nor that trembling could never be an action (though we might speak of "pretending to tremble") nor does it imply that a bodily movement could not be involved both in an action and a reaction. We often speak of a child crying in order to gain attention, but this does not, I think, mean that we regard the crying as wholly a sham, as not an emotional reaction at all. Perhaps what the child does as an action is to refrain from refraining from crying whereas its crying is a reaction rather than an action. (49)

There is a further difficulty for our analysis with what behaviour scientists call displacement behaviour - in response to the frustration of a desire an agent might engage in some apparently unrelated activity. If a victim of unrequited love breaks down and cries we might call this purely a reaction but if he wrings his hands, buries his head in his hands etc. it seems that he is performing actions, though it is not clear just what his

reasons are which is not to say that it is unclear which desire is at the root of his behaviour. On my analysis these do not seem to be actions at all. They are not performed with a view to some further end, in particular with a view to satisfying his desire for the woman, and if we postulate desires to wring his hands, bury his head in his hands etc. this can seem a little ad hoc and leaves the connection of these desires with the desire for the woman unexplained. I think this difficulty will be encountered by any theory of the distinction between action and reaction - for any account of displacement behaviour must satisfy three conditions. Firstly, it must represent the behaviour as a sequence of actions. Secondly, it must connect the behaviour with the frustrated desire. Thirdly, it must not represent the behaviour as purposive, that is engaged in with a view to some further end, in particular to satisfying the frustrated desire. And these conditions together seem to yield the conclusion that an action can be performed because (in some fairly immediate sense) of a desire even though the action does not satisfy the desire nor is believed to bring about its satisfaction, and this not because the action involves some kind of mistake like turning right when one wants to turn left. So any account of the distinction which interests us must cater for action which are intelligible in a way which does not see them as appropriate to bringing about the object of the desire from which they spring.

I think we should be struck by this problem - that an action can be intelligible as an action performed because of a desire even though it seems irrelevant (even given the agent's beliefs) to the coming about of the object of that desire. And an account of the distinction between action and reaction must come to terms with the fact that some bodily movements which occur because of a desire are not considered actions, whereas others are, even though they exhibit the teleological irrelevance we are considering. Let me make a suggestion: some desires when frustrated (or fears where one sees no way of avoiding what is feared) lead to desires to perform certain actions in a way that is not intelligible in means-end terms ; this is in itself evidence that a person has these desires, whereas if the desires to which they led were intelligible in means-end terms this intelligibility would be dependent on the assumption that the person had certain beliefs. This may make seem some of the phenomena we have been considering as reactions more like actions (for example, one can say, "I wanted to cry") but otherwise it seems intuitively acceptable,



If someone is prevented from wringing his hands or burying his head in his hands when he suffers some bereavement, he may avow that he wants to do so. This suggests that the introduction of desires in this type of case is not idle. So we can treat the occurrence of certain desires (<sup>or</sup> in actions assuming these desires are unimpeded) as non-contingent evidence for the occurrence of certain other desires. The evidence is non-contingent in that it depends on our concept of the desire and its natural manifestations, and not on the person's having some matching belief in a particular case. On the other hand, it must be admitted that our concept of such desires would not be very different if such behaviour were not general, though our concept of people's emotional life might be.

If someone has a desire or intention, then sometimes he will be aware of it - it will become the object of his critical attention. By this I do not mean merely that it is conscious but that the person can take up attitudes to his own desire or intention. One might want not to have one of one's desires, or one might feel ashamed of one's intentions and if one can have attitudes one can act on these attitudes or they can lead to bodily reactions. Thus someone might take action to rid himself of a certain desire (such as seeing a psychoanalyst) or his shame at a certain intention might cause him to blush. Thus the possibility arises that the original intention or desire to which one has an attitude causes <sup>the</sup> ~~a~~ bodily movement involved in the action or the bodily reaction. A certain desire or intention might cause a certain attitude to it which causes a bodily movement which is either an emotional reaction to one's having the desire or intention or is an action for which the attitude is a reason. This possibility presents a difficulty for our causal analysis of action and the distinction we made between action and reaction above.

Frankfurt in an article (as far as I know unpublished to date) discusses the following thesis:-

"M is an action of which P is the agent if and only if M is a movement of P's body and there is some description of M under which 'The occurrence of M is explained causally by the fact that P intended that M should occur' is true."

Against this he makes the following point - "there is an obvious difficulty, however, in this way of explicating the notion of an intentional bodily movement. Suppose I am at a party and

have decided to spill what is in my cup. By prearrangement this will signal my confederates to draw their guns and kidnap our host and hostess. But when I lift my cup from the table on which I have laid it, and have the fateful instrument actually in my hand, the horrific import of the events I am about to initiate becomes vivid to me. Second thoughts begin to race through my mind and my anxiety rises. Then panic ensues, bringing with it the involuntary shaking and trembling of my hands which causes the coffee in the cup to spill. In these circumstances, surely, spilling what is in the cup is not something I do intentionally. Neither are the spasmodic movements of my hands, from which the spilling results. Yet it may be that I intended to spill the cup by making movements of just these kinds and that the movements are causally explained by the fact that I intended that they should occur. Thus a movement may be a mere happening in a person's history even though it occurs because the person intended it to occur."

What we seem to have here is an intention causing an emotional reaction to itself, the emotional reaction involving a bodily movement. There may be more than one way in which this could occur or its occurrence analysed - for example, it might be the conflict between the person's intention and his attitude towards it that causes the movement, or it may be attitude that causes the movement or again it may be the intention which causes the bodily movement as a reaction to it, this being part of what it is for the person to have that attitude towards it. But what seems to ensure that Frankfurt's case is a counter-example to the proposed definition of an action of which P is the agent, is that an intention can cause a psychological reaction to itself, and such a reaction might involve a bodily movement of just such a kind as the person intended to perform. And clearly Frankfurt's example could be modified by substituting "desire" for "intention" so as to produce a difficulty for our analysis of action. (50)

The possibility of intentions causing psychological attitudes to themselves can also be used to produce examples of bodily movements which although intentional are not intentional in virtue of certain of the intentions which can truly be said to have caused them. Thus if, in the above example, the man, on becoming horrified at his intention, had acted on his horror in some way, say by rushing out of the room, the intention could be said to be a causal factor leading to the bodily movements involved in rushing out of the room but could not be the intention in virtue of which

the bodily movements were intentional. This has a certain relevance to the problem of displacement behaviour - here we have bodily movements which although intentional are not intentional in virtue of certain of the desires which can be truly said to have caused them. If the frustration of a desire leads to certain actions of this kind then we should say that the actions were performed because of the desire but not in a sense which would give the intention with which the actions were performed. The main difference between displacement behaviour and reactions to one's desires and intentions seems to be that in the former case the reaction is to the frustration of the desire, whereas in the latter it is to the existence of the desire. (There may be the same teleological irrelevance in the latter case as in the former).

To return to the difficulty posed by Frankfurt's case - it is not, of course, an attack upon the causal theory of action, for it in no way tends to show that causation by a desire or intention is not a necessary condition of a bodily movement's being intentional, but it does show that there is going to be some difficulty in giving sufficient conditions for intentional action. Earlier I suggested that if a desire causes a bodily movement this is an action if the bodily movement is of the type specified by the object of the desire, or if the bodily movement is believed by the agent to be such as to bring about the object of the desire and this belief plays a causal role in the production of the bodily movement. Clearly this does not deal with the Frankfurt type of case. (51)

Cases of this type will not, of course, be common. They depend on there being certain bodily movements which are characteristic of emotional states and which are not then action but which could also be performed intentionally. But it seems likely that cases could be constructed which do not involve this kind of mechanism involving mediation by emotional states. Suppose it were known that someone desired to raise his arm; this knowledge might lead someone else to raise his arm for him. It could well be that the man's desire to raise his arm was a causal factor leading to his arm's going up (especially if one subscribes to a causal theory of knowledge). And yet we would not say that he had raised his arm. Clearly, one wants to say that his arm's going up must come about by the right kind of causal pathway; to adapt the well-known slogan from the theory of knowledge one might say that a man raises his arm if it is no accident that his desire to raise his arm (or that his arm should go up) is satisfied. But this does

not tell us in what the right kind of causal pathway consists, or when we are to speak of "no accident". We can specify a number of wrong causal pathways, so we might be able to say that in certain circumstances a desire's causing a bodily movement constitutes an action provided that it is not the case that... and here we mention a number of conditions that would defeat the ascription of action. Unfortunately, I do not have a way of giving a complete list of such defeating conditions and of demonstrating its completeness; and even if I did, one might feel that what is needed is a positive characterisation of what it is for a bodily movement's being caused by a desire to constitute an action. One feels a certain inclination to characterise the Frankfurt type of causal pathway as somehow indirect and to think of those involved in ordinary action as somehow direct. But as an analytical suggestion this seems to lead nowhere, as presumably even in the case of intentional action the causing of the bodily movement by the desire is mediated - by events in the nervous system and the muscles and for some reason these are permissible.

Now the man in Frankfurt's example would normally know he had not performed an action. And if he had performed an action, that is if he had carried out his plan to spill the contents of his cup deliberately, he would have known this as well. What is more, he would know these things non-observationally (52) and without appeal to neurophysiology. He might know nothing of the neurophysiological processes involved in action or emotional reaction, and yet be able to say something that amounted to claiming that either the right or wrong kind of causal process had occurred for this bodily movement to be intentional. We could not, I think, use his sincere belief assertion as a decisive criterion of whether the movement was intentional. For consider the following case: we know that someone desires to raise his arm and by means of some electrical wiring of the muscles of his arm, we bring it about that his arm goes up. And let us say his desire to raise his arm is a causal factor in leading us to do this. Then even if he sincerely claimed to have raised his arm, we would override this assertion and regard the bodily movement as not intentional. And yet a person's sincere avowal as to the status of a bodily movement of his as intentional otherwise does seem to be basic to our ascription of intentionality to a person's bodily movements. We could not regard all cases of apparently intentional bodily movements as like the above, as then the distinction between intentional and unintentional bodily movements would collapse.

Perhaps the notion of an intentional bodily movement is best explicated by reference to a paradigm: a desire to perform an intentional bodily movement causes that bodily movement without the mediation of emotional states or external states of affairs (such as manipulation by an outside agency). In such a case an agent will normally, if he is sincere, admit that he performed an action. That this correlation holds is essential to our concept of action (in a sense it is not a mere correlation). The reason why certain causal pathways are excluded can be explained by reference to our concepts of self-determination, autonomy and the sovereignty of an agent's desire. The satisfaction of a person's desire in the case of outside interference is, so to speak, by the grace of the manipulation<sup>or</sup> and depends on his desires and in the case of emotional mediation as in the Frankfurt example runs counter to one of the attitudes that causes it. The point of desires is that the world should be made to correspond with them - in our deviant cases the correspondence occurs but not in a way that suggests the likelihood of future correspondence and in Frankfurt's case actually involves the frustration of one of the desires or attitudes which brings about the bodily movement (the prospect of trembling and its consequences is just what horrifies the man).

It is conceivable that neurophysiological discoveries could lead us to make small-scale revisions to our demarcation of intentional action. If it were discovered that a small number of what seemed to be intentional actions (ie. none of our defeating conditions were satisfied) involved a radically different causal chain from desire to satisfaction, we might come to view such cases with suspicion as perhaps not real cases of action. But no large scale revision would be countenanced without abandoning the concept of intentional action as we know it. The concept is built on two struts - what we have called the point of desire and an agent's awareness of the significance of his own behaviour. In the end, it looks as though one will have to analyse the concept of action in a way that involves the untidy notion of defeasibility (53) as only then can the deviant causal pathways be explained as deviant - as out of step with the significance of desire. One just has to account for them piecemeal.

Finally, I want to discuss another type of action which does not fit in well with an analysis so far. This is the type of action which is not obviously intentional under any description,

such as idly drumming one's fingers. Here one is not drumming one's fingers because one believes that drumming them will bring about some further end, nor is it clear that one is drumming them simply because one wants to. It is tempting to make the move made earlier and consider someone who is idly drumming his fingers and is then prevented from doing so. He might experience some dissatisfaction and this would suggest that he had been frustrated in his desire to drum his fingers. But he need not do so, and even if he does, his dissatisfaction might be more a matter of resentment against interference. In the case of displacement behaviour, we are dealing with a wide-spread pattern of response to certain situations and one feels that in all or most cases a person indulging in such behaviour would feel further frustration if prevented from doing so. This seems normal, and if someone did not do so, one might just regard him as odd or accept a similar oddity in any explanation of his behaviour - such as that his desire to bury his head in his hands had that moment ceased. Or one might be so persuaded of the existence of such a desire, as to say that perhaps because of the man's emotional state he was simply not conscious of it. But there might be a residue of displacement actions which are only with artificiality accommodated to this view, and in any case it does not seem plausible as regards idly drumming one's fingers.

If someone is idly drumming his fingers and is aware of this, one might consider his action of drumming his fingers as intentional, but be unhappy about talk of any desire or other pro-attitude towards drumming his fingers. If he is not even aware that he is drumming his fingers, even the idea that his drumming his fingers is intentional would be cast in doubt. In the former case, although one might feel that his drumming his fingers was intentional, one might still not want to speak of his intention to drum his fingers. All one seems to be saying is that he accepted that he was drumming his fingers (as when one speaks of an accepted consequence of an action as an intended consequence). What makes this type of action a difficulty for us is the absence of anything which is anything like a desire to drum one's fingers. And there is the problem for any analysis of action, as to how one is to distinguish this drumming one's fingers from one's fingers drumming without one's drumming them.

Someone who is unconsciously drumming his fingers could presumably notice he is drumming them, or better, realise he is drumming them. (54) Someone whose fingers were drumming without

his drumming them could certainly notice they were drumming. In the two cases one might suppose that his reaction would be different. But this cannot be simply explained in terms of surprise. As Strawson (55) points out, someone who regularly experienced involuntary bodily movements might be quite unsurprised by their occurrence. And someone might be surprised to find that he had been unconsciously drumming his fingers. If his fingers were drumming without his drumming them, one would want to say that he somehow lacked control over the movement of his fingers. But a natural interpretation of this might easily be too strong for he might be able to stop his fingers moving. True he would stop his fingers moving, rather than stop moving his fingers but this assumes the distinction we are trying to explicate.

If his fingers were moving without his moving them, would he have to exert effort to stop them moving? Possibly like stopping himself trembling (where we can speak of some part of his body or of his whole body trembling; one does not have to say that he is trembling). Suppose someone's fingers are drumming, and as soon as he notices this, they stop drumming. He has no time to stop them drumming. It seems dubious to say that he does not know whether he was drumming his fingers or his fingers were drumming because he had no time to find out whether he stopped drumming them or stopped them drumming (against resistance).

We might tend to assume that if someone is idly drumming his fingers, then there will be some explanation of this which likens it to intentional behaviour, for example, an explanation in terms of some kind of nervous tension which the drumming relieves. The trouble is that an element of artificiality appears once one begins to speak of the agent drumming his fingers in order to relieve the tension. And talk of unconscious or preconscious desires to reduce the tension seems theory-laden; such concepts are of relatively recent origin whereas I doubt very much whether the recognition of unconscious actions such as drumming one's fingers unconsciously is at all new. Still, unless one thinks of the drumming of the fingers as somehow having a point it is obscure why we should think of it as an action at all. I said earlier that there might not be an absolutely clear cut distinction between action and emotional reaction, because of an unclarity as to the applicability of teleological notions in some cases. This unclarity seems to introduce a similar vagueness into the distinction between an action and a mere bodily movement.

If an animal freezes at the sight of danger, one might be equally inclined to describe this as an action as to describe it as an emotional reaction. Pressed to define the former usage, one might have to admit that one had little warrant for ascribing relevant beliefs to animals. With a lowly animal one might even be unhappy about ascribing fear to it. And yet the idiom of action does not seem out of place; we talk of instinctive and reflex actions. This shows, I think, that we use the idioms of action and of teleology (we might talk, for example, of the function which freezing serves in animals of that sort) even when the paradigm conditions for their applicability are far from satisfied. Our readiness to describe someone's unconsciously drumming his fingers as an action can be similarly underpinned by such ideas as that drumming them relieves some kind of tension. One can say that drumming one's fingers relieves nervous tension and that that is why one drums them. And one might even say that the purpose of drumming one's fingers is to relieve nervous tension. But one would not say that one drums one's fingers with a view to reducing nervous tension. The idea that behaviour can be seen as significant and analogous to consciously purposive behaviour even when the appropriate desires and beliefs are absent gives us a middle-ground between mere bodily movements and fully-fledged intentional actions.

These considerations serve to remove some of the puzzlement about the very existence of actions which seem to spring from no obvious desire or pro-attitude, and yet which we are reluctant to demote to being mere bodily movements. <sup>They do</sup> It does not however accommodate such cases to our analysis of action so far. And indeed it is fairly clear that we have, in Wittgenstein's words, been nourishing ourselves on a one-sided diet. Not all actions, it seems, can be analysed in terms of the causation of bodily movements by desires and the like, although perhaps those that cannot/<sup>can</sup> always be rendered intelligible by comparison with parallel cases which we might construct that do involve desire and sometimes belief. Obviously, where human beings are concerned, these cases could not be the only ones. And if one focusses exclusively upon human action, behaviour such as drumming one's fingers and reflex actions are bound to seem degenerate cases of action. Our analysis of action is essentially one of action for a reason and of intentional action; and I shall not attempt to supplement it by an analysis of the sorts of quasi-teleological behaviour which we have uncovered, behaviour which does not appear to involve desires and the like.



One other point ought to be mentioned, however, a point which might help to lift these actions out of the limbo to which a too rigid insistence on the dichotomy between actions and mere bodily movements threatens to relegate them. When someone performs an action, the bodily movement involved will have many features which were in no sense intended or desired by the person in question. If someone lights a cigarette, the exact details of the trajectory of his arms is not likely to be of significance to him. And yet these are features of an intentional action. Now someone can intentionally sit down and intentionally remain seated. Sitting in the sense of being seated is hardly an action (though refraining from getting up in some cases is) though it can be intentional; one might be sitting because one wanted to do and still does (or "still wants to remain seated" if one rides hard the maxim that one cannot want what one believes one has already got). Many features of one's posture will not be intended or desired and yet they are all aspects of one's state of being seated which is intentional. Now one might be drumming one's fingers whilst sitting and perhaps this too can be regarded as an aspect of one's intentional state of being seated. Just as an unintentional action of one sort is describable as an action in that it is intentional under some description perhaps one's drumming one's fingers might be describable as an action because one's total state is intentional under some description. (56, 57) This involves stressing the unity of, for example, a person's posture at a certain time and if someone wanted to deny that the drumming of a person's fingers was an action, it would place the burden of proof on him to show that the unity was somehow broken in this case. This, however, is a mere suggestion; it may involve some illicit juggling with the notions of the descriptions of an action or state and the aspects or features of an action or state. It is of interest however in that it does not depend on the assumption that the drumming of one's fingers has, in a broad sense, a point. It might appeal to someone who felt that even the introduction of such things as nervous tension was misplaced a priorism.

Chapter 5

Intentions : For the Future

I have argued that if someone performs an action, and his reason for performing that action is some desire of his, then this desire causes the bodily movement which is involved in the action. This account makes no mention of intention and this is clearly a serious omission. Intentions are certainly involved in action, but how? Let us approach the question by asking what the relation is between desire and intention. If someone performs an action intentionally then according to Nagel it is appropriate to ascribe to him a desire to perform that action. But if someone performs an action (his reason for performing it being a certain desire) it is appropriate to ascribe to him an intention to perform that action under a description which links it with the desire in one of the two ways discussed. This would seem to make the connection between desire and intention very intimate indeed. But this takes no account of intentions to perform actions in the future, actions indeed which one may never in fact perform. Not only that, but by leaving one kind of intention out in the cold, it might seem to bifurcate the notion of intention completely. One can avoid giving this impression by pointing out that intentional actions of some complexity can be seen to involve intentions for the future if one imagines that someone performing such an action were stopped in the middle of it - he would normally be able to say how he intended to go on. Furthermore, many intentional actions are examples of preparatory behaviour; they are performed with a view to making it possible for one to do something else later on.

These considerations show that the problematic type of intention is found lurking in the types of intention about which we thought we had some understanding. They do not help us much with that <sup>2</sup>action itself. In the case of intentions for the future, although from a knowledge of someone's intentions one might be able to deduce something about his desires (that he has a desire to perform the action that he intends to perform, or some other kind of reason to perform the action) one cannot from a mere knowledge of a person's desires deduce anything about his intentions. This suggests that we investigate how the transition from desire to intention is <sup>e</sup>affected. I shall begin by considering certain differences between desires and intentions.

(a) One can desire a state of affairs, and the specification of the state of affairs need involve no reference to one's own action. One can also intend that a state of affairs should come about and the specification of that state of affairs again need involve no reference to one's own actions. But here there is a difference: to desire that something should be the case does not involve one's even contemplating an action, whereas to intend that something should be the case is to intend to perform some action to bring it about, or to refrain from performing some action which would prevent it. Intentions, one could say, are always intentions to do. A strong distinction between a desire and an idle wish might seem to weaken this contrast, but to destroy it completely it would have to be too strong to be plausible. A desire that is entirely overruled by countervailing considerations may lead to no contemplation of action. It does not thereby cease to be a desire.

(b) One can have a desire and also have a desire which conflicts with it - we might be able to satisfy either one of the desires, but perhaps because of some feature of the existing situation or perhaps because the objects of the desires are logically incompatible, one cannot satisfy both. To have conflicting desires is not to stand convicted of any kind of irrationality. The position is different with intentions: one can indeed have conflicting intentions (as when one makes appointments which clash) but here one cannot avoid the charge of irrationality or at least error.

(c) It is hard to see how someone who intended what he believed to be impossible could avoid the charge of irrationality. If this seems too strong (Anscombe and Thalberg would probably say that it was (58) ) let me put it another way: if one intends what one believes impossible, whether one's belief is false or one's intention is doomed to frustration, but more than that, for either one's belief is false or one is mistaken about what one is going to do. Whereas, if one desires what one believes impossible, one's belief may be a mistake but if it is not, although one's desire is frustrated, one is not thereby convicted of some kind of error. Intending what one believes impossible is bound to involve error. Desiring what one believes impossible is not. If one insists that a desire for what one believes impossible is really only an idle wish, one can reformulate the point in terms of the genus constituted out of desires and wishes. An idle wish can lead to the same bodily and emotional reactions as a desire. Furthermore, it seems that given this distinction between desires and idle wishes, desires can become idle wishes and vice versa merely as one's beliefs change. One might object to my claim that if someone intends what he believes

impossible either his belief is false or he is wrong about what he is going to do. To make such an objection would be to deny that entailment from "I intend to do X" to "I am going to do X". As far as I can see this has some plausibility provided one does allow an entailment to "I am going to try to do X". If one does not, one seems to break the connection between intentions for the future and the future and to make avowals of intention into avowals merely of one's present mental state. If, on the other hand, one does allow the latter entailment, then at least one's intention to do something is linked with what one believes possible (or does not believe impossible) in that one must believe it possible to try to do (or not believe it impossible to try to do so). "I am going to try to do X and I believe it impossible to try to do X" seems radically incoherent.

Property (b), that of the uniqueness of intentions, seems to be one that desires cannot have. Whereas desires can be desires to perform actions but need not be, and also can be desires for what one believes possible but need not be, it does not seem that any desire could be such as to rule out the possibility of a conflicting desire even given that the agent is behaving rationally. Conflict seems to be a permanent possibility within one's system of desires. New beliefs could always bring two desires between which there was formerly no apparent conflict, into conflict. This might, of course, happen with intentions but they could not both be rationally maintained as intentions once one had accepted the fact of conflict. This consideration seems to militate against regarding intentions as a species of desire: the uniqueness cannot be built into the specification of the desire. It may happen not to conflict with any other desire one has, but this is an extrinsic fact about it. It is not contingent that when someone announces his intention this creates a presumption that he has no conflicting intentions.

Some animals (perhaps most) are not capable of forming intentions for the future. In such animals there seems no real distinction between intention and desires. For intention is only manifested in action and here we ascribe intention and desire *pari passu*. If it is asked: how else, even in human beings, is intention manifested but in action? I would accept this provided action is made to include verbal avowals but would point out that human beings know of their own intentions without having to go on their manifestation in action (if one does like the word "know", one can speak of "awareness" or say that human beings can be in a position to avow

their intentions even if they do not). At least in animals the distinction between desires and intentions is of a piece with the distinction between desires that lead to action and desires that do not. I said earlier that the uniqueness property prevents us from simply regarding intentions<sup>For the future</sup> as desires. These considerations about animals show why it might seem plausible to maintain that intentions are, or are reducible to, desires. Another reason one might have for thinking that intentions are desires is the point made by Anscombe that the reasons are given for one's intentions are the same sort of reasons as are given for one's desires and not the sort of reasons one gives for predictions of one's future behaviour (59) (I do not, of course, wish to imply that Miss Anscombe reduces intention to desire).

One trouble with our account of the difference between desire and intention is that it is difficult to give an account of how an intention can possess this property of uniqueness, in possessing which it differs from a desire. We can formulate the problem in Humean terms - how can the existence of an intention as a matter of logical rather than contingent fact create a presumption that something else namely a conflicting intention, does not exist? On one level it is easy to give an answer: if someone claimed to have two intentions which he recognised as incompatible and refused to give either of them up or admit the necessity to give one up or even to re-examine the facts to see if the intentions could be reconciled, we would say he failed to understand the concept of intention. We might think he was really talking of desire. But this does not go to the heart of the matter.

We could create a concept of a P which is that of a desire which is not opposed by other desires. And we could make the same point about a person not understanding the meaning of having a P as we did about his not understanding the meaning of intending. But having a P would not be the same as intending. Having a P would not be a necessary condition of intending as one can intend to do something for one reason, which one also has a desire not to do. Nor is having a P a sufficient condition, as this does not imply anything about one's believing the object of one's P possible or about his believing that anything he does will have any bearing upon whether the object of his P will come about. If someone desires X and this desire is unopposed by any other and he believes that X is possible and either X is "to perform an action" which he believes he can perform, or an object which he believes he can get,

or a state of affairs which he believes he can bring about or a state of affairs which he believes he could prevent from coming about, does he necessarily intend X? The answer is unclear but I think that it is probably in the negative briefly because to accept the entailment would be to leave no room for the notion of forming an intention or at least to make it an empty ceremony. On the other hand it could be said that it is in just such cases of unopposed desires that there is no need and hence no room for deliberation.

At any rate it is more important to stress the fact that having a P is not a necessary condition of intending. It is of the greatest significance that one can intend to do what one has a desire not to do, or what will prevent the satisfaction of other desires that one has. It would be artificial to maintain that the desires which emerge as the losers when one forms an intention disappear and in any case this would be to put the cart before the horse as the desire which wins is opposed by other desires to begin with and yet it is possible to form an intention in this situation. So we have (i) that an intention, as a matter of non-contingent fact, cannot in general co-exist with an intention one believes conflicts with it; but (ii) a<sup>2</sup> intention can, and after deliberation typically does, co-exist with desires that one believes conflict with it.

Whereas I have argued that reasons for action other than desires are most conveniently bracketted with desire, it seems that there are good reasons for keeping the concept of intention for the future separate from desire. Another argument that an intention for the future cannot be a desire runs as follows:- if we say that the intention is one of the desires involved in deliberation or conflict, the one that proves triumphant after deliberation(it could hardly be some other desire) we must specify some property which it has which marks it as an intention. This cannot be that it is not opposed by other desires as it is. It cannot be that it is not opposed by other intentions as this involves the concept we are trying to analyse. It cannot be that it is the one that leads to action as it may not or if we allow (wrongly I think) action to include the formation of an intention this again assumes the ~~analysandum~~ and what is more separates it from the desire.

There is another suggestion that cannot be eliminated so easily. Perhaps the intention is the desire which the agent

can non-inductively predict will move him to act. One might prefer not to say that the intention is the desire, but say: a person intends to do X if and only if he desires to do X (where the desire to do X might be a desire for a means) and can non-inductively predict that this desire will move him to act. There are difficulties with this idea, but first I would like to point out its merits. Firstly, it seems to explain the uniqueness of intentions, for one cannot rationally make two conflicting predictions. (60) Secondly, it deals with conditional statements of intentions: roughly, if someone says he intends to do something if P, then he is non-inductively saying that the desire to do that thing will move him to act if P. Thirdly, it explains the connection of intention with action. Fourthly, it does not suggest that the formation of an intention is an action (I shall explain why I do not think that the formation of an intention is an action later). And fifthly, it shows why the reasons one gives for what one intends are of the same type as the reasons one gives why one wants something or why one does something and of a different type from the reasons one gives for a belief. For if one non-inductively predicts one does not offer evidence but the prediction concerns a desire which one can attempt to justify and this is the right kind of justification for intention.

The difficulties in this view of intention are three in number. Firstly, do we say that a person who intends to do X, can or does non-inductively predict which desire will move him to act? If the former, what justifies the prediction? - the only answer seems to be that he has an intention. If the latter, in what sense is this true if he does not avow his intention? Secondly, in my definition I allow that the desire for X might only be a desire for a means. The trouble is that it is not clear that when someone desires Y, believes that X is a way of getting Y, he necessarily desires Y even when he acts so as to get Y with a view to getting X. (61) Desires for means often seem somewhat shadowy cases of desires. If one tries to get round this by talking of the desire for the end, one cannot then deal with intentions which involve a choice between alternative means. Thirdly, might someone simply have an intuition of his future behaviour? He makes a prediction of his future behaviour of which desire will move him to act, he makes it non-inductively but it is not an intention. (62) It may of course not be right, but then one can say what one intends to do or what one intends should be the case and then one might change one's mind. This difficulty merges with the first one in that one is apt to say that, in the case of

an intuition of one's future action one is not entitled to make the prediction. So in a sense one cannot make it. It diverges from the first difficulty in that one's intuitions might be justified - or at least one might be justified in accepting them because they had repeatedly proved correct. Of course, this type of justification would rule out intention; but if someone just has an intuition of his future behaviour, we can ask what justification he has for his belief. The answer is 'None', and how are we to distinguish this from intention? (63)

If it is said that someone who intends to do X simply justifiably believes that he will do X without having a way of knowing this, we can ask what entitles him to this belief. Either we appeal to the concept of intention, or we say he just knows of his entitlement without having a way of knowing. Then we can ask what entitles him to the belief that he is so entitled; after all someone with an intuition of his future behaviour could feel entitled to his belief. And so the regress continues. Furthermore, this analysis of intention does not seem to deal adequately with the concept of deliberation. If deliberation were simply a process whereby one arrived at a belief, this would either involve an assimilation of deliberation to theoretical reasoning or would be unintelligible, a mental charade which is terminated but not consummated by the appearance of a belief.

One point I have slurred over: intuitions of one's future behaviour are by definition without justification, although acceptance of them might have a justification if, for example, they had repeatedly proved correct, but then this alone would rule out intention. Intentions, on the other hand, do have a justification, a justification of the same type as one gives for desires or actions. And this does not seem to suggest the difference we need? The trouble is that it is not clear how this is to be integrated into our theory. If someone's having certain beliefs about a certain desire moving him to action is a necessary and sufficient condition for his having a certain intention, provided that the intention is justified in a certain way, we are obviously involved in circularity. So how should the definition be modified? We cannot say that it is the desire that is justified in this way, as this might apply to any desire and we can hardly say that it is the belief as how can a belief be justified in this way? So we seem to be saying that someone has a certain intention if and only if he has a certain belief about a desire of his, and this belief is justified in virtue of the justification of his desire. But this,



apart from its obscurity, will obviously not do: for the reasons one gives for an intention often involve considerations and weighing of all one's desires not just of the desire which <sup>wi</sup> runs out. To this extent justification of one's intentions differs from justification of one's desires, although it does not differ from justification of one's actions.

At this point, if not earlier, one is likely to be struck by the resemblance between intentions ~~to~~ perform actions and actions. The justification one gives for either present or past intentions (often excuses) seem exactly similar to those for either present or past actions (again, often excuses). Both intentions and actions seem to have a certain uniqueness - just as one cannot rationally have conflicting intentions one cannot equally perform incompatible actions. The intentions, it seems, foreshadow the action: one may not be able to act immediately but one can at least form the intention to act. The intention seems almost a substitute for the act. One is prevented from acting now, or the action one contemplates has, for some reason, to be at a later date so one intends to perform it instead. These analogies are compelling. There is, however, one important error to which they might lead, namely the idea that the formation of an intention is itself an action. One feels inclined to view the formation of an intention as an action because of the role it seems to play as a kind of provisional substitute for action. On the other hand, it is hard to see how the formation of an intention could itself be unintentional or intentional. One can perhaps, of course, intentionally engage in deliberation with a view to forming an intention but this does not make the formation of the intention intentional. And it does not seem that the formation of an intention is necessarily intentional in the way that murder is intentional killing in certain circumstances. For what is the X-ing that forming an intention is intentionally doing? Nor is it necessarily unintentional in the way that making a mistake is necessarily unintentional. For what kind of acts could one be attempting or performing which could be described as forming an intention because of some un<sup>and</sup>interpreted character or consequence of what we did?

Let us take a simple case of someone forming an intention. He desires to do A and perhaps has opposing desires; later on he forms an intention to do A. We can ask why he formed the intention to do A. And, as we have seen, we would expect the same type of answer as we would expect if we asked why someone intentionally

did A. The important point here is that we would not expect any special reasons for forming the intention to do A, reason for thinking the intention to do A desirable. And this, I think, shows why it is wrong to think of forming an intention as an act. One would justify forming the intention by justifying the object of the intention, not by characterising the intention or the forming of the intention as desirable. In this respect the forming of an intention and the forming of a desire for a means are similar. If one desires Y, believes that X is a means to Y and for this reason comes to desire Y this is not an action (not that there is much temptation to think that it is in this case). And the reason is that justifying the desire for X would involve saying what was desirable about X, in this case that it was a means to Y, and not saying what was desirable about desiring X.

Let us now return to the difficulty in which we were enmeshed when trying to state necessary and sufficient conditions for someone's having an intention. We accepted Anscombe's point that intentions are marked out by the type of justification one gives for them. But then it became obscure, when attempting to analyse intention, what part of the supposed analysis was going to be the part to which this kind of justification was appropriate. We noticed earlier that analysing intention in terms of predictions about one's desires was rendered somewhat implausible by the fact that it was not clear when one could ascribe a desire for a means to someone. Now, as a characteristic way of stating one's intentions is to say "I am going to..." which sounds like a prediction and makes no mention of desire, it seems natural to drop this talk of desire and think of intentions as, in some sense, predictions about what is going to happen or what one is going to do.

In an earlier chapter, I defended the coherence of the notion of bifunctionality. What was clearly a belief both in its form and in the justification appropriate to it might nevertheless, it seemed, motivate action, like a desire but without the need for some further desire to make this possible. It is natural to try to apply the notion of bifunctionality in our analysis of intention. For supposing someone expresses his intention by saying "I am going to do A". This is an indicative statement about the future and has the same form as that in which a predictive belief about the future might be expressed. It also seems subject to falsification by future events in that if I do not do A it seems that my statement "I am going to do A" was false. On the other hand the justification

one offers for the statement of intention is not what is likely to happen (although it would involve what is possible should happen) but what it would be good to make happen. And furthermore, if one says one is going to do A, but does not, and this alone constitutes a mistake then the mistake is in the performance, not, or at least not only, in the remark. (64)

Here we seem to have something analogous to bifunctionality; intentions have two directions of fit. In our earlier discussions of bifunctionality it was possible to say that certain beliefs could function like desires. There is no comparably simple formulation of the situation with intentions. Intentions are like *désires* in the justification one offers for them, but in this respect they are even more like actions. They are also like beliefs about the future in that in avowing an intention one asserts that something is going to be the case. But it is not obvious how to exploit this fact in an analysis of intention. It is beginning to look as though intentions are *sui generis* - not that they cannot fruitfully be compared to desires and beliefs: the point just made does just this; but that they cannot be reduced to them or regarded as some kind of species of desire or belief.

I may have given the impression that intentions to do something in the future do not by definition lead to action so long as they remain intentions for the future. This, of course, is untrue for they can lead to preparatory behaviour, actions performed in order to put oneself in a position to do something else later. This kind of intention for the future would not present quite the same difficulties, for here we can speak of someone's intention in acting, and perhaps treat this as the desire that is moving him to act, in this case the desire to do something else later or to be in a position to do it. Once again we see why it can seem natural to treat intentions as a species of desire. But this still leaves us with the problem of someone's having an intention which is not yet moving him to act and which perhaps never will, but which is nonetheless a genuine intention.

An idea that springs to mind at this point is that such an intention, if it is genuine, would move one to act if one were in a position to act now or the time were right. When the time comes one might not act but then one must have changed one's mind - one cannot intend to do something at *t*, and this intention persist until one believes that the moment *t* has arrived, and one recognise that the time at which one intended to act has come and yet not act unless

one at that moment changes one's mind. In Analysis Problem No.7 the question, "Can I decide to do something immediately without trying to do it immediately?" is discussed by Brian Ellis, "Candidus" and Nicholas Rescher. (65) There seems to be a fair amount of agreement between these three contributors so firstly I shall endeavour to summarise their conclusions.

The contributors seem to agree on the following:-

- (a) There are cases of deciding to do something immediately without trying to do it immediately
- (b) If I decide to do something immediately, I cannot fail to try to do it immediately because I change my mind (Ellis) "The point of the problem-setters 'immediately' is to exclude the move of avoiding the verdict that I have failed to comply with my decision by suggesting that I must have changed my mind. I cannot (logically) revoke my decision at the very same time that I decide (Candidus). Rescher, on the other hand, mentions change of heart "as an interpretation of" I decided to do thus-and-so immediately but did not immediately try to do it." He does not give an example so it is not altogether clear whether he would really wish to disagree with the other contributors.
- (c) Some of the examples of deciding to do something immediately and failing to try to do it immediately, depend on the first "immediately" meaning "as soon as possible" and not with no (perhaps psychologically discernible) time-lapse whatsoever. Ellis's example would seem to be of this kind. This is a perfectly defensible usage but perhaps does not deal with the most important problem. Also it tends to involve us in the somewhat difficult problem of distinguishing the preparatory moves directed towards putting oneself in a position to do something and attempts to do that something. (65a)
- (d) A man may fail to try to do immediately what he has decided to do immediately because he is unable to try or something interrupts him so that he forgets his decision. A man's mission is completely forgotten when he notices his fountain-pen - it is forgotten when a loud explosion occurs outside (Ellis) "I found myself unable to do it", "My courage failed me", "I couldn't even try", (Candidus) - "While talking to him by telephone, I decided to tell him off, but he hung up" - "I decided to call out his name, but would not bring myself to do it" - "I decided to tell the password, but did not remember it" (Rescher).

Can one decide to do something immediately without trying

to do it immediately, where one does not have to do anything first in order to put oneself in a position to do it or try to do it. This reformulation is intended to exclude (c) but it also excludes some of the other cases eg. "I decided to tell the password but did not remember it". Here what one has to do first is something mental and as such has the peculiarity noted by Rescher. One might prefer to think of this as something having to occur before one is in a position to do what one has decided to do. So let us add to our question the qualification: "and where nothing has to occur before one is in a position to do it". Another case was the following:- "While talking to him by telephone, I decided to tell him off, but he hung up". Here the man does not have to do anything nor must anything occur before he can tell off his acquaintance, but certain of the conditions must remain the same, that is certain things must not happen, otherwise he will find himself unable to carry out the telling off. Let us therefore add another qualification to our question: "where there is nothing that must not happen if he is to be in a position to do it or to try to do it." Actually, this formulation is somewhat stringent as there maybe no situations in which a man's decision to do something could not go unaccompanied by a try because of some upsetting event, for example, the man's death. The reason why I have formulated it in the way I have is that the qualifications then seem to exemplify a pattern. They read:

- (a) where he does not have to do anything first in order to put himself in a position to do it or to try to do it
- (b) where there is nothing that must not happen if he is to be in a position to do it or to try to do it
- (c) where nothing has to happen first that will put him in a position to do it or to try to do it.

Now about what seems to be a missing element in the pattern:-

- (d) where there is nothing that he must not do if he is to be in a position to do it or to try to do it? Could someone decide to do something immediately but fail to do it immediately because he did something immediately he made the decision which left him no longer in a position to make the attempt. Hardly if what he did was intended to have this effect. This would be revoking his decision at the very same time as making it - or making a decision in the face of a contrary decision which one had not revoked. But in so far as one can do two things at once, it seems that one could unintentionally foul one's attempt to implement one's decision. By fiddling with the telephone, our phone caller could cut himself off.

What other ways of failing to implement one's decision are left? - "I decided to call out his name but could not bring myself to do it", "I found myself unable to do it", "My courage failed me". These are presumably cases of what Rescher calls "impotence". The argument that "Candidus" present for their existence is interesting - "Consider these two examples (a) 'I decided to hit him immediately he entered the room' (b) 'When he entered the room, I decided to hit him immediately'. Only (b), of course, exemplifies our problem. But if I may fail to comply with decision (a) not because I have revoked it, but because at the crucial moment 'I found myself unable to do it' or 'my courage failed me', as may I fail to comply with decision (b). To say 'I tried but I couldn't' would be misleading: it would suggest physical handicaps or obstacles. And there is no need to fill the gap with psychological handicaps or obstacles. It is not that I tried and could not. I could not even try."

Now this brings us round to the problem we are interested in: can one intend to do something at  $t$ , and this intention persist until one believes that the moment  $t$  has arrived, and one recognise that the time at which one intended to act has come, and yet not act? And it would seem that the answer is, "Yes" - the same kind of examples as Ellis, "Candidus" and Rescher present, could be adduced to demonstrate this. Furthermore, it seems one could also change one's mind at the moment one thought that  $t$  had arrived. None of these would seem to affect the status of the previous intention as genuine. Can we salvage some logical principle to the effect that if someone intends to do  $A$  at  $t$ , and when he believes that  $t$  has come he does not act, then he must have changed his mind or something (in the widest sense) must prevent him? And if so would such a principle help us with our analysis of intention, or would it be reduced to triviality by its lenient escape clauses?

We do at any rate seem to have reached some point of contact between intentions to act in the future and intentions in acting. But the difficulties in taking this analysis further are obvious:-  
(a) We have not dealt with intentions to do something some time or other. To do so we would have to drop the reference to the person's realising that the time at which he intended to act has come, and instead speak of his realising that an opportunity has arrived, and that there are no countervailing considerations etc.  
(b) This last modification has a familiar tag. In the first chapter, I inveighed against a supposed entailment from desire to action.

The new entailment from intention to action sounds similar, and it may be doubted whether it will fare any better. In particular, it looks as though one will have to countenance the possibility of someone's being wrong to think he has a certain intention. And is this allowed for by our concept of intention any less grudgingly than someone's being mistaken in thinking he has a certain desire is allowed for by our concept of desire?

(c) As I said above, the leniency of the escape clauses which we might have to allow in order to produce a logical truth might deprive it of usefulness in any analysis. Furthermore, there is a strong suggestion of circularity in the provision of the clause "or he must have changed his mind", for this involves at the least the revocation of the previous intention, and, possibly, the formulation of a new one.

In fact, I think that at one level the charge of circularity can be met. Suppose we have a principle of the form, "If someone intends to do A at t, believes that t has arrived, and recognises that this is the moment at which he intended to do A, then either he does A or tries to or p (a disjunction of the possibilities that Ellis, "Candidus" and Rescher allow) or he revokes his intention". This seems plausible because if he does not act, then provided not-p, his not acting itself constitutes a revocation of his intention to do A at t (he might, of course, still intend to do A at some later time). On one side of the connective, we have a proposition about someone's intending to do A at t at a time earlier than t, whereas on the other we have a proposition about someone's abandoning his intention to do A at t at t, so there is no real danger of circularity. Unfortunately, this does not show that the principle could without circularity be used in a definition of intention. For here, the fact that the word "intention" occurs in both sides of the connective would alone present a difficulty. Furthermore, our principle seems to suggest only a necessary condition for someone's having an intention.

Suppose we take as basic the notion of an unrevoked intention. It is a necessary condition for someone's having an unrevoked intention to do A at t, that when he comes to believe that the time (t) at which he intended to do A has arrived, he either does A or tries to do A or p. I do not think that this logical truth can be refuted in the same way as the supposed entailment from desire to action. This is because we are taking failure to act in certain circumstances not as a criterion that

that someone did not have a certain intention, but as a criterion that he must no longer have it when he fails to act. We are not allowing the present to determine as a matter of entailment what the past must have been like. We do not have an equivalent concept of revoking a desire, though we do have a concept of losing a desire. Might we not treat inaction in certain circumstances as a criterion that someone had lost a certain desire? I doubt it, because our tendency to say "In failing to act, he ipso facto revoked his intention" seems to derive its compulsiveness from the fact that we are here regarding his failure to act as intentional, and we are unwilling to allow conflicting intentions. There does not seem to be any equivalent source for the idea that someone who fails to act in certain circumstances must necessarily have lost the relevant desire. If one wanted to adopt it, it would have to be treated as plausible in its own right.

Our idea of taking the notion of an unrevoked intention as basic does not seem to lead to a definition of an unrevoked intention (in terms of which we could perhaps define a revoked intention). For we still have only a necessary condition for someone's having an unrevoked intention, and moreover there is circularity apparent in the condition that the person must believe that the time at which he intended to do A has arrived. At this point, I feel forced to abandon the attempt to analyse intention for the future in the sense of providing a necessary and sufficient condition for someone's having such an intention, a condition which does not itself involve deploying the concept of intention. It seems that any proposed analysis is bound to seem reductive, although in some cases its reductiveness might only be demonstrable by the construction of somewhat <sup>real</sup> ~~erondite~~ cases, for example, the possibility of "intuitions" of one's future behaviour. (66) There seem to be quite a number of close conceptual connections between intentions for the future on the one hand and desires, beliefs and intentional actions on the other and it may still be possible on the basis of these to construct an analysis of intentions for the future, a reduction of such intentions to these concepts. But I do not see, at the moment, how this can be done. So I shall proceed by assuming that intentions for the future are, in the last resort, irreducible and return to the problem stated at the beginning of this chapter.



Chapter 6

Intentions : as Involved in Action

Someone has a number of conflicting desires; on the basis of one of them, or of a consideration of them all, he forms an intention to perform A in the future. I have argued that we should not think of this as an action. For if we ask for his reasons for forming the intention, we get reasons for performing A. The agent has no reasons for forming an intention over and above his reasons for what he intends to do. (67) Someone might want to say that this merely shows that one's reasons for forming an intention are of a special sort, unlike one's reasons for forming a society, where one could mention the merits of the society or of one's forming it, but this does not show that forming an intention is not an action. But then it becomes unclear what the resemblance is supposed to be between forming an intention and other kinds of action. If forming an intention is an action, then it looks as though, although one has a reason for performing the action, the reason is neither some merit one sees in the action itself, nor some consequence of it. (It would be very odd, for example, for someone to say, "I want X. If I form the intention to get X, I will probably act on it. So I will form the intention to get X, because this intention is a means to my getting X.") Furthermore, if one were to allow that performing an intention was a special kind of mental action, the reasons for performing which were not features or consequences of the act or the intention, but the same reasons as one has for the object of intention (what one intended) it would be difficult to see how one would block a similar move as regards conceiving a desire for X because one believed it was a means to Y which one desired; ~~was~~ a special kind of act the reasons for which were not merits of having or forming the desire, but of the object of the desire. And we should not want to say that conceiving a desire was an action (whereas at least in the case of forming an intention, the word "forming" does suggest some kind of action). I am not saying that if someone desires Y, and believes that X is a means to Y, he must always desire X, nor even that if his desire for Y moves him to act so as to get X, this must always occur via the generation of a desire for X. But, clearly, one desire's leading to another desire (which we should have no hesitation in describing as a desire and not merely the having of a reason for doing something) by means-end reasoning is a process which sometimes does occur.

Finally, someone might want to say that forming an intention is an act for which one does not have a reason in the usual sense. This seems obscure unless it means that forming an intention is an action which is necessarily unintentional, like making a mistake (under that description). But what conceivable analogy is supposed to hold here? Or with reflex, automatic or involuntary actions? I think this suggestion can safely be left in obscurity.

Let us therefore return to our agent who has a number of conflicting desires who on the basis of one of them or of a consideration of all of them, forms an intention. The reason for the disjunction is that I do not wish to imply that the formation of an intention must depend on a balanced assessment of all the relevant considerations, that the formation of intentions implies some special degree of judgement. If one is to defend a causal theory of action, then it seems that one must find a way of fitting intentions into this causal nexus. The most obvious way is along the following lines: the desire to do A at t causes an intention to do A at t which at t causes the appropriate bodily movement. Now this is to attempt to fit intentions for the future into one's analysis of intentional action. A possible alternative might be to leave such intentions out of the causal chain leading to the bodily movement: one might be inclined to do this if one regarded an intention as some kind of belief about one's future behaviour. The intention itself would have to be accounted for causally (consider: what is the force of the "because" in "I intend to visit him because I want to ask his advice") but it could be on a causal chain presumably branching out from the desire, but other than the one leading to the bodily movement.

In what sense are intentions involved in action? An action can be intentional: someone can have a certain intention in acting: someone can have an intention to do something in the future and when he believes the time has come he does it: someone, because he intends to do something X at a later date, does Y in order to be in a position to do X at that later date: someone, because he intends to do X at a later date, does Y because of some anticipated consequences of his doing X. In all these cases an intention seems to be involved in an action. The last two cases are both examples of preparatory behaviour, but there is a significant difference between them. Suppose someone intends to spend a certain sum of money on a car at a future date; he realises that this will leave him short of money for other purposes and so starts to save

frugally now. It would be artificial to describe him as saving money in order to be able to spend it on a car, for he intends to buy a car irrespective of his success in living frugally (perhaps he would be prepared to go into debt to do it); rather he is saving money in order to offset the consequences of his buying a car, a course of action on which he has already made a prior decision. The situation is not particularly different from taking action in order to deal with consequences of someone else's action because one knows of his intentions, except, of course, as regards the question as to how the intentions are known. On the other hand, if the man is saving money in order to buy a car (or be able to buy a car) there is a more intimate connection between his saving money and his intention to buy a car. It is noteworthy that in the first case there is a temptation to speak of his knowledge of his own intentions and certainly one would want to speak of his knowledge of what he was going to do, but these locutions are unnatural when one is speaking of someone doing something in order to put himself in a position to do something else.

The importance of this is that we have found a clear case of intention being involved in action in the sense that one's knowledge of what one is going to do is involved in one's action. In the other examples, this is not obviously true. Against this we have to weigh the facts that in these other cases one's knowledge of what one is doing may be involved in the action in that if one were stopped in the middle of one's action <sup>one</sup> ~~we~~ would normally be able to say how one intended to proceed, and that if one is doing something in order to be in a position to do something else, one's knowledge of what one is going to do if one is successful in putting oneself in a position to do it seems to be involved.

At this point, the direction of my argument may become clearer if I state what I hope to establish. Briefly, I want to show that motivation can be explained in terms of desire ( and related concepts such as fear etc.) causing bodily movements, and in so far as one has to recognise an irreducible concept of intention this is essentially involved in motivation in one case only, that of acting in order to offset certain consequences of one's intended actions. And in this last case it is the knowledge of (or belief about) one's future behaviour - whatever else there may be to the concept of intention - that is important.

Suppose it is said that intention plays an essential part in action, and by this is meant that an analysis of action cannot

be given which does not involve the irreducible concept of intention. Consider someone who intentionally performs an action for a certain reason. His reason is a certain desire, and we say that his desire's causing the relevant bodily movement constitutes his performing the action. Has anything been left out? We saw earlier that characterising intentional actions in this way presented difficulties in distinguishing them from certain bodily reactions. The most difficult problem came from a case suggested by Frankfurt. Now, here it should be noted that Frankfurt's problem was posed in terms of the inadequacy of a definition of intentional action in terms of an intention's causing a bodily movement. So it rather looks, though if anything has been left out of our attempt to characterise intentional action, which omission makes it unable to deal with the Frankfurt-type case, <sup>this is</sup> not simply ~~in~~ reference to the causal role of intentions, as the problem will still arise if we include this.

So far we have mainly been considering intentions as prior states and enquiring whether it is necessary to give them a causal role in actions. But this may be misleading: it amounts to restricting intentions to intentions to act in the future and these, we have seen, are not the only kind of intentions. Suppose someone performs an intentional action - we can inquire with what intention he performed the action and we can inquire in virtue of what was his action intentional? But here it is not clear that we are asking questions which cannot be posed or answered without reference to intentions! For "With what intention was the action performed?" we can ask "Why did he do it?" or "For what reason did he do it?" It would be very odd to say "I want to know his intention in doing that, not his reason for doing it," or "I want to know his reason for doing that, not his intention in doing it." Nevertheless, I think in certain cases, sense could be given to both these locutions. Foreexample, if asking why X fired a gun on Y, we received the answer "To kill him" we might produce the latter locution. But equally we might say, "I know (or assume) that his reason for firing the gun at Y was that he wanted to kill him: what I want to know is why he wanted to kill him". And if watching a chess-match, we ask why someone made a certain move and receive the answer, "In order to win", we might just possibly produce the former locution, but equally we could say, "I assume his reason for making that move was that he wanted to win; what I want to know is what reason he had for thinking that move would enable him to win", or "I assume he intends to win; what I want to know is how he intends to win with that move."

Perhaps, behind this, there lurks some principle about one's tendency to speak of intentions when speaking of the contemplated immediate effects of our actions and reasons when these effects are at further remove (in both cases one must speak of the intended effects, but equally one can speak of the desired effects) (68) However, what I hope these considerations show is that there is no fundamental difference between asking for a person's reasons for acting and his intentions in acting. (69) And this removes one source of the inclination one might feel to regard intentions as well as desires as causally involved in action.

Consider the following case: someone has a certain number of conflicting desires. On the basis of a consideration of these, he forms an intention to perform a certain action at time t in the future. When he believes that t has come, his intention causes the relevant bodily movement. Now it would seem that the man has performed the action he intended to perform. The trouble is that he might, in the mean time, have lost the desires on the basis of which he formed his original intention. So whatever reason we give for his action, it cannot simply be his desire or desires which led him to form the original intention. (70) This suggests that if an action is to be performed for a reason, that reason must be causally involved in its genesis in a way which depends upon its still being present when one acts. It will not do for a desire to generate an intention, and then cease to exist, whilst the intention lives on to cause the relevant bodily movement. Nor will it do for the desire to cause an intention which later causes a bodily movement, if at the time one has forgotten one's reasons for forming the intention (even though one may still have the desires in question) at least as long as the desire is not unconsciously operative.

We have thus two kinds of reasons for thinking that the role of intentions in acting is the same as that of reasons - if intentions are causes, this is to say no more than that reasons are causes. There are several objections we must face at this point:-

- (a) Where does this leave "intentions to act in the future"?
- (b) Where does it leave "doing something/<sup>X</sup>because one intends to do something else Y later, where X is intended to offset the consequences of one's doing Y"?
- (c) If someone is stopped in the middle of some action, he will normally be able to say how he intended to go on. Yet this is

neither giving a reason for some action, nor does it seem uninvolved in the action, a mere contingent fact about it.

Normally intentions to act in the future, on our view, are not in themselves involved in action, and we have seen that to allow them to play a causal role in the genesis of actions, distinct from that of the reasons, is conceivable, but means that the action might be performed for a different reason from that for which the intention was performed. In normal cases of action where one does what one intended earlier to do, this is not the case. And then there is no reason to suppose that the intention for the future has its own causal role. Objection (b) derives its plausibility from the fact that we can in such cases say, "He did X because he intended to do Y later" and this cannot be translated as "He did X in order to do Y later". We seem to have a case of an intention's explaining an action though not by giving a reason for it. But, of course, this appearance is only superficial. This intention explains the action only if we assume certain desires on the part of the agent; ex hypothesi this intention does not itself give the desire. This intention explains the action in the same way as beliefs normally explain actions. A belief about his future actions leads him to expect a certain situation which is relevant to his desires and it is this that explains his action.

Objection (c) must be explained more fully: it does not seem to be merely a contingent fact that when I am stopped in the middle of some action I can usually explain how I intended to go on. This was giving an intention to act in the future and yet seems bound up with what I was doing. Now, in such cases how I intended to go on would have had a special relevance to what I was doing. Thus, if at a certain stage in some connected sequence of actions say A, I am stopped and asked how I intended to go on, then my answer may "To perform B" would be related to A in a variety of possible ways - I was going A because only then could I do B, I was doing A because otherwise there would be no point in doing B, I was doing A because I was carrying out some ritualistic procedure in which A happened to be followed by B. Knowledge of how one intended to go on is intimately related to knowledge of why one was doing what one was doing when one was stopped - otherwise why think of the person as being engaged in a piece of connected behaviour at all? For it is also possible for someone to be doing something without the least idea what he will do when he has finished it. Or to intend to do something completely unrelated afterwards.

I hope that by now I have given some plausibility to the view that one does not have to prepare a special place for intentions as distinct from reasons in the analysis of intentional action. Where intentions come into their own, as clearly distinct from reasons for actions, is in the case of intentions to act in the future. Why, it may be asked, should we use the same word in both cases? We can ask a man what his intentions are, and we can ask him with what intention he did something. These are two uses of "intention" which in some ways seem poles apart. The latter, we have argued, is a request for a reason: the former is clearly no such thing. Now to both questions the answer one gets is likely to mention some envisaged future state of affairs. In the case of asking someone for his intentions, this seems necessary; but in the case of asking someone for his intention in doing something it is not - "What was your intention in raising your arm?" - "To signal". (71)

So it does not seem that the various uses of "intention" can be unified by saying that they all refer to the future. This is certainly true of the word "intentional", as when one says that a bodily movement is intentional. Earlier I mentioned a property of intentions for the future which I called their uniqueness. This can be seen as a kind of commitment to a course of action. When one acts with a certain intention, some desire or desires will have in a limited way proved triumphant; and so it is when one forms an intention to do something. It is this, I think, that explains why we use the word "intention" in such seemingly diverse cases. One is referring to the fact that a person progressed beyond the stage of merely having certain desires - some of them have resulted in action or commitment to a course of action. Although it is wrong to think of the formation of an intention as itself an action, it is natural to think of the formation of an intention as foreshadowing the act which is intended. To speak of intention is to attribute some kind of efficacy to a desire. There are one or two possible exceptions. Sometimes one says that someone intended something to occur, meaning that he anticipated it as a consequence of his action and accepted it, although he in no sense desired it. But I think it is fair to call this a loose usage simply because such attributions of intention are often rebutted by giving the fuller description of the state of affairs - "No, I did not intend it; I merely accepted it as a consequence of my action" (72) To say that such examples are best kept apart from intention proper is not, of course, to say that if someone does sincerely and correctly deny having a certain intention as above, he thereby absolves himself of responsibility for the

consequence of his action. Another case is that of someone's consciously drumming his fingers - one might want to say that he is intentionally drumming his fingers without wishing to imply that he has any desire to do so. But here again usage is not clear cut, and one might prefer to say merely that he is consciously but not intentionally drumming his fingers.

My main objective above has been not to produce a complete theory of intention for the future but to show that my analysis of the motion of the intention with which something is done does not produce an unbridgeable gulf between this kind of intention and <sup>that</sup> of intention for the future.



## Chapter 7

### Mental Action

So far our analysis has dealt only with actions involving overt bodily movements. It has had nothing to say on the subject of omissions, refrainings and mental acts. A natural move would be to treat such actions as the causing by desires etc. of certain bodily states and mental events. Doubtless, this concept will play a part in a final analysis. But in itself it sets no limits to the possibilities of human action. In a sense, this was true of the causal analysis of actions involving bodily movements. However, there is a difference between questions like, "Why cannot people move their hair at will?" and "Why cannot people believe something at will?". It seems that the question as to what bodily movements one can perform at will is to a large extent a contingent matter, (73) (it becomes less obviously so when one asks why someone cannot move a table without doing anything which brings about the motion of the table, although even here it might be possible to maintain the contingency thesis by claiming that the limits of one's body are contingent). But as regards mental actions, it often seems a necessary truth that one cannot perform them at will, or perhaps it is better to say that in many putative cases there can be no such actions.

There are many acceptable English phrases which suggest that desiring and believing are things we do, things we can decide to do or try to do. For example, "I decided to accept his word", "I decided to reject that hypothesis", "I tried to shut the fact out of my mind", "I decided I wanted a gun", "I tried to stop wanting it". Such locutions might incline us to the view that belief and desire are or can be subject to the will, that is believing and desiring are things we can do or try to do or refrain from doing at will. I want to oppose this view. Before developing a systematic critique, it is worth considering why the examples given do not really support such a view. (74)

Take, "I decided to accept his word". To accept someone's word is perhaps not always a matter of believing him at all; it might simply signify that one refrains from an outright accusation of dishonesty out of politeness or because one cannot prove him a liar. When it is a matter of belief, the deciding to accept his word is deciding to believe him, but one cannot decide to believe him without believing him. One can believe him without deciding to

believe him; this would happen when no process of weighing up his story or his character occurred. But deciding to believe him does not seem to be anything over and above coming to the conclusion that his statement is reliable. Deciding to believe, it would seem, is not like deciding to act, where one's decision might go unimplemented. This does not show that one's desires are not in some way responsible for what one believes. But it does remove the temptation to think of "deciding to believe" as akin to "deciding to act", and hence involving one's desires in just the same way, and to just the same extent.

A similar treatment of "I decided to reject that hypothesis" can be given: I cannot decide to reject a hypothesis without deciding not to believe it, and without it being, in some sense, true that I come not to believe it, and in considering the process of coming to the conclusion that something is not to be believed, there is no point where it seems one needs to introduce a desire. The case of "I decided I wanted a gun" is not entirely dissimilar. If it is to appear even a *prima facie* case of desires subject to the will, it must not be taken as a judgement that one has a desire (of which one was previously perhaps only dimly aware) but rather as reporting the conclusion of a piece of practical reasoning- "I wanted to shoot him, So I decided I wanted a gun". Here again there is no gap between deciding one wants a gun and wanting a gun. On the other hand, it is a desire (wanting to shoot someone) that leads to the desire for the gun. But this original desire is not a desire to desire a gun, so again one's decision is not like deciding to act. I am not sure how much significance to attach to the fact that one says, "I decided I wanted a gun" and not "I decided to want a gun", even when one is reporting not the discovery or recognition of a desire, but the conclusion of a piece of practical reasoning. But this fact alone might deter one from regarding "I decided I wanted a gun" as indicating that wanting is something one can do at will.

The two remaining examples, "I tried to shut the fact out of my mind", and "I tried to stop wanting it" do not seem amenable to the above treatment. They do not express decisions and cannot be interpreted as the conclusions of pieces of theoretical or practical reasoning. Perhaps one should begin by looking at the corresponding statements, "I decided to shut the fact out of my mind", and "I decided to stop wanting it". These are significantly different from the sorts of decision discussed above; whereas one cannot decide to believe something without believing it, one can

certainly decide to shut something out of one's mind and yet fail to do so, and one can also decide not to want something and yet still want it. (75) In fact what seems hard to conceive is that a person should decide to shut something out of his mind, and as simply shut it out of his mind in that someone should decide to cease desiring something and as simply refrain from desiring it. (76)

Clearly people can manipulate their beliefs and desires but the examples that spring to mind seem to involve doing something to bring it about that they believe, do not believe, desire or do not desire something. Can we call this believing or desiring at will? I think that such a terminology would tend to obscure important differences between these cases and those where it is more obviously correct to describe someone as doing something at will. Let us agree that someone can bring it about that he desires that p somehow or other (say, by getting someone to hypnotise him); let us also agree that he can do this intentionally. We are considering the legitimacy of collapsing "He can intentionally bring it about that he desires that p" into "He can desire p at will". Now, of course, one use of "at will" is precisely to exclude the need for means and to refer to cases where one does something without <sup>adopting</sup> having means to do it. But this may not be the only use. There are many sentences of the form "He can intentionally bring it about that p" which entail sentences of the form "He can X"; for example, "He can intentionally bring it about that John dies", entails "He can intentionally kill John". Here, in passing from the antecedent to the consequent, we introduce a new verb. What we are looking for is a case where we can pass from "He can bring it about that he X's if he wants to" to "He can X if he wants to". Possibly, we can move from "He can bring it about that he dies if he wants to" to "He can die if he wants to". So, maybe to allow an inference from "He can bring it about that he desires (or believes) p if he wants to" to "He can desire (or believe) p if he wants to" would not extend any special indulgence to desire and belief. However one thing one would not do in the above cases would be to describe his dying, believing or desiring as intentional (one would describe them as intended). Nor could one describe them as actions. And it seems that the clearest way to bring this out is to use the expression "He can intentionally bring it about that he desires p" rather than "He can desire p if he wants to" (77)

So far we have argued that, although a man is not powerless to alter his beliefs or desires, those cases where he intentionally affects his beliefs or desires should not be characterised as

desiring or believing at will. But why cannot one desire or believe at will? A natural answer as regards belief is that it is the point of belief to match the world as it is, and the ability to have the beliefs one wanted would interfere with the function. This does explain how it is that one can have a measure of long-range control over one's beliefs (78) ?/not nor does it suggest an explanation of why it is that one cannot desire at will. To take the latter point, we cannot say that desire would lose its essence if we could desire what we wanted to desire, for it is not clear that there is some conceptual connection that would be broken by this, in the way that belief at will would break the connection between belief and the world. And yet there seems to be something wrong with desires which appear in this way.

As a first shot, we might ask: would not such desires be divorced from the needs of the agent? Suppose that someone desires to desire that p. It is important that he does not already desire that p, that we do not mean by his desiring that p that a desire for p, which he has already, should be stronger or actually move him to act. Before considering his ability to bring it about that he desires that p, let us ask: why does he desire to desire that p? Frankfurt gives an example in which a reason is given why a narcotic's doctor desires to desire heroin - because he believes his ability to understand his patient would be increased if he knew what it was like for him to desire the drug (79). Suppose, in this case, that his second order desire is satisfied and he comes to desire heroin. We might ask why he desires heroin. Now on the assumption that he desires heroin solely because he desired to desire heroin, certain answers are ruled out (trivially). He does not desire heroin because heroin is a means to some further end. This would make the efficacy of his second order desire irrelevant. And if the only reason that he desires heroin is that he desired to desire heroin then even if he did believe the heroin to be a means to some further end, this would not be the reason why he desired it.

Thus it seems that if people were able to desire at will, the desires they obtained thus would be adventitious. I hope it is clear that the fact that the desire for p is connected with the desire to desire that p, in that it results from it and accords with its object is irrelevant. The object of the adventitious desire stands in no such relations as that of means to end to any of his other desires, or, if it does, then only by accident. Actually it may be doubted whether Frankfurt's example is really a good one for our purposes. For the reason why the doctor wants to desire heroin is that he wants to know what it is like. And surely the only way he can be certain of

finding this out is by inducing in himself the kind of desire for heroin that addicts normally have. And to do this he would have to induce in himself the desire in the "normal" way, by taking heroin until he has become addicted. This would be long-range manipulation. But it is hard to think of a plausible case of desiring to desire which suits our purposes better. So it looks as though simple desire at will is going to be odd, not only because of the adventitious character of the first order desires that result, but because of the peculiarity of the second order desires that lead to them. Frankfurt regards even his own example as standing "at the margin of preciosity".

However the oddity of the desires involved may not be a conclusive reason for rejecting the idea of desire at will. Let us try another line of attack, and in doing so, bring the question back into relation with the possibility of belief at will. Suppose someone were to want to want *p* (or to believe *p*). On this basis he decides to want *p* (or believe *p*). And - lo and behold! - we find he now wants *p* (or believes *p*). The usual behavioural criteria for his desiring (or believing *p*) turn out to be satisfied. Obviously we would not be satisfied with this alone as an example of desiring (or believing) at will. Indeed there is a rather large battery of questions we should want to ask.

How do we know that he did not previously want (or believe) *p*? How do we know that it is because of his desire to want (or believe) *p*, that he wants (or believes) *p*? And if it is because of his prior desire to want (or believe) *p*, that he now wants (or believes) *p*, does this make it an action? Could he, particularly as regards belief, have full awareness of his power and of its having been brought to bear in a particular case? Did he, prior to his supposed "mental act", have any beliefs which might serve as a basis for his believing *p*, or have any desires which might lead to a desire for *p* by some process of practical reasoning? Could he cease to desire (or believe) that *p* if he wanted to, and if so, is he aware of having this power? How does he justify to himself his desiring (or believing) *p*? And in the case of some beliefs that *p*, how is his belief related to his perceptions? Does he ignore them, or is he able to reconcile his beliefs that *p* with them? Or are even his perceptions under his control? Here we would find it difficult to avoid assimilating belief at will to fantasy, and perception at will to imagination. And, in the case of some beliefs that *p*, how is this belief related to his other beliefs? This last point is somewhat different in that it might just conceivably be made

---

out to support the possibility of belief at will - one thinks of the claim that any proposition can be held to be true provided we are prepared to make a sufficient number of alterations to the rest of our beliefs. (80)

But, in fact, it affords an illustration of what is wrong with the idea of belief at will in any simple form. For if a corollary of the fact that it is the point of belief to match the world is that, if one finds one has been entertaining contradictory or contrary beliefs, one has to relinquish one of them or admit the necessity for this. Otherwise one cannot be said to believe one proposition rather than the other (perhaps it will be said that one believes neither). But if one could believe at will, one could hold beliefs that were contrary. Unless that is, we start applying restrictions to the effect that we can only have a belief at will when it has no contrary beliefs. But we cannot get even this far by such a line of reasoning when we come to desire. For there is nothing unacceptable about the idea of conflicting desires.

From the welter of questions adduced above, let us see if we can discover any real incoherence in the idea of desire at will. The man who claimed to desire to desire that p, and then claimed that because of this he desired that p raised the questions: how do we know that he did not already desire that p when he claimed merely to desire to desire that p, and how do we know that he did not already have desires that might lead to a desire that p by a process of practical reasoning. No doubt we can stipulate, ex hypothesi, that the man does not already have such desires but could we ever be sure in an actual case? It seems a conceptual possibility that someone might have a desire to desire to raise his arm (but no desire at the time to raise his arm) and that this could cause a desire to raise his arm. The second order desire is very odd, perhaps a psychologist of an introspectionist school wants to investigate such a desire in himself, let us say a burning desire to raise his arm. Suppose his desire for a burning desire to raise his arm causes the appearance of a burning desire to raise his arm. I shall assume that we can refer to some precipitating<sup>ing</sup> event (perhaps, only the second order desire's reaching a certain intensity) so that we are not enmeshed in difficulties about how one state can cause another.

I am not sure about the plausibility of the example I have chosen; it may be that I am implicitly taking over some of the assumptions inherent in introspectionist psychology itself. However,

what I want to argue is that, and this seems generalisable to all putative cases of a desire to desire that p causing a desire that p, our conceptual system surrounding desire protects us in a large measure against having to accept the existence of any actual cases of this. For what would show that the person who claimed to desire to raise his arm because he previously desired to desire to raise his arm, was right? Surely we would want to be certain that he now actually desired to raise his arm, and the best evidence for this would be his actually raising his arm. Admittedly, we have argued that a desire even if unopposed need not necessarily lead to action, but in such a peculiar case as this we would be unhappy about accepting the man's claim unless we had incontrovertible evidence that he now desired to raise his arm. But then, if he did raise his arm, this would cast doubt on his first claim that he had previously only desired to desire to raise his arm. Was it really true that he had no desire to raise his arm, but desired to desire that he raise his arm? Or was it that he already desired to raise his arm and also desired that the desire to raise his arm should move him to act? To establish a case of someone's desiring to desire p causing a desire for p one must be sure that both desires exist at the right times and not at others, and be exactly/as described. One seems compelled to place absolute reliance on the person's testimony if one is to accept a case of desire at will. (81)

Whereas, in considering belief at will, it can be shown that subjection of a putative belief to the will tends to destroy one's entitlement to regard it as really belief (82), there is no equivalent argument as regards desire at will. So far, we have had to content ourselves with indicating both the oddity of the resulting desire and of the second order desire from which it springs; and with showing how conceptual considerations about desire and action seem to insulate us from being faced with an incontestable case of desire at will. The argument on the latter subject has involved a distinction between straightforwardly second order desires, desires to have a certain desire which one does not already have, and desires which are also in a sense second order, desires that certain of the desires one already has should actually move one to act. Frankfurt calls the latter 'second order volitions'. A digression on this subject is in order.

The statement "I raised my arm" is not well analysed as "I caused my arm to go up". One argument against this analysis (83) is that if the action is to be intentional then my causing (of my

arm to go up) ought to be an intentional action, and therefore ought to be analysable in the same way. But it is unclear how this is to be done, and even if it were, it would seem to generate a regress, and what is more, a regress possibly involving the somewhat obscure notion of "causing to cause". Our theory of action in terms of the causation of bodily movements by desires etc. does not, of course, involve us in this unacceptable analysis. However, it might be that there is a sub-class of actions which are of the form "I caused some movement of my body", for example, if I lift my right arm with my left arm. Not only can one not analyse "I raised my arm" as "I caused my arm to go up", one cannot analyse "I raised my arm" as "I caused myself to raise my arm". Regressive tendencies are here even more apparent. But again it might be maintained there is a sub-class of actions which are of the form "I caused myself to perform an action" (84). For example, if I want to ensure that I perform a certain action at some future date, say because I doubt my own strength of character, I might arrange matters so that when the time comes the only possible or reasonable course of action open to me is the action I want to get myself to perform. The explorer who burns his boats so that he has to go on to explore the hinterland might be an example of this.

What the explorer is doing here is ruling out the possibility of satisfying any countervailing desire, so that the only desire on which he would be likely to act is the desire on which he wants himself to act. And one might, though perhaps only with some latitude, describe this as causing his own actions. Some at least of the examples of this kind seem to involve a certain degeneracy. For example, if I get myself to perform a certain action A by giving myself a drug which makes the desire to perform A overwhelming, when it would otherwise not be, then it is not implausible to say that I somehow undermine the status of A (let us say that it is not a free action). One would be even more inclined to say this if one got someone else to perform an action by administering a drug or some such means.

It might be said that the reason why the explorer can cause himself to explore the hinterland tomorrow by burning his boats today, is that he does allow himself a certain freedom tomorrow; he could remain on the beach and starve. The reason one can cause an act, is that one leaves a certain latitude, a certain freedom to choose some other act. This view is likely to be held along with the view that one can only cause actions in a somewhat loose sense.



One can express this somewhat existentialist standpoint that an action can never be completely and utterly compelled, that one can never strictly say, "I did it because I had no choice" as follows: if one fully determines a future action of oneself or someone else leaving no leeway or room for manoeuvre, no room even to make an unreasonable choice, one ipso facto deprives the action of the status of a full-blooded free action. This may be true on some interpretation (85) but what I wish to enquire here is: what happens if one attempts to close the temporal gap between one's present action and the future action it is supposed to cause?

One is immediately conscious of a certain oddity in this suggestion, particularly if it is formulated as a precise analogue to the case discussed above: someone at  $t$  causing himself to perform an action at  $t$ . For this reason, it is worth enquiring why its conceivability or otherwise might become a problem for someone. One is occasionally faced with the following sort of dilemma: it might well be within my power to do something, but is it within my power to do it for the right reasons? I might want to do something for moral reasons, but I might also want to do it to avoid incurring censure or to obtain praise. It may not be possible for me to do it without its coming to the notice of others, but I want to act on the moral considerations rather than out of regard for public opinion. Can I act on the desire or value on which I want to act? Obviously, I might just happen to act on the desire on which I want to act, if, for example, I forget the other desire (assuming that it does not unconsciously influence me). But can I ensure that I act on the desire on which I want to act? And if I must act now can I ensure now that I act on the desire on which I want to act?

It does not have to be the case that the rival desires must lead to the same action, as above. I might be torn between desires to do different things. But for some reason or other, possibly a moral reason or because I identify with one of the desires, I want one of the desires to prove effective. I must act now; can I ensure now that I act on the favoured desire? (86) If this is conceivable, it seems, at first sight, as though one has two actions here for one has two desires: the desire that leads one to perform a certain action, and the desire that this desire should be effective in leading to that action. Alternatively and more plausibly, perhaps one only really has one action to which both desires are, in different ways, contributory. Frankfurt speaks of someone who exercises freedom of the will as "securing the

conformity of his will to his second order volition" meaning by this not merely that he is moved to act by the desire that he wants to move him, but that it is because of his second order volition that he acts on the first order desire (A second order volition, on his terminology, is a desire that a certain desire should move one to act, and one's will is the desire that does move one to act.) It is not merely a happy chance that his will conforms to his second order volition. Leaving aside the question as to whether Frankfurt is speaking of the freedom of the will as usually conceived, it is clear that he is concerned with the problem of the efficacy of second order volitions, which is the problem which interests us. (87)

There are several possible positions on this issue. One could deny that second order volitions ever are effective, except via the long-range manipulations discussed earlier (the example of the explorer). One could deny that the effectiveness of my desire that a certain desire should be effective and the effectiveness of the embedded desire are distinct. They together lead to the action I perform. This is basically the position that we have here only one action to which both the second order volition and the first order desire are contributory (88). On this view securing the conformity of one's will to one's second order volition would be acting on both the second order volition and the favoured first order desire. Finally one could deny that there is any distinction between the effectiveness and the satisfaction of a second order volition. For most desires there is such a distinction: if I want to be in Paris, this desire is effective if it leads me to take the action directed towards going to Paris, even if I do not even get to Paris, whereas it is satisfied if I end up in Paris, even if I do not do anything to bring this about. If this distinction is inapplicable to second order volitions, then it is inappropriate to ask whether my second order volition is involved in its own satisfaction. The trouble with this is that wants to distinguish the satisfaction of my second order volition by outside interference (say, someone, without my connivance, administering a drug to me) and the case where my second order volition has something to do with its own satisfaction. It seems that this position really amounts to the first one, the denial that second order volitions ever are effective. And this would still be true if one distinguished between the accidental satisfaction of my second order volition (by involving such things as outside influences) and its satisfaction because of the natural unmanipulated strength of my first order desire.

So I want to investigate whether anything can be done with the second position, the idea that both the first order desire and the second order volition can contribute to action. The question must therefore be answered as to how these two desires with different objects can reinforce each other. Now one obvious way in which two desires with different objects can reinforce each other arises if their objects are related as means to end. Is this applicable to the present case? The object of the first order desire is X say, whereas the object of the second order volition is that the desire for X should move one to act. Clearly the object of the first order desire is not a means to the object of the second order volition. Is the object of the second order volition a means to the object of the first order desire? It seems that one must normally believe this to be so: if one believes that one can in a certain situation act on a desire one must believe that one's action is a means to satisfying that desire. In some cases one will have to speak not of means and ends but of something being a way of satisfying a certain desire. If I desire to raise my hand and desire that the desire to raise my hand should move me to act, then, in normal cases, the efficacy of the desire to raise my hand will consist in my raising my hand (I say "in normal cases" because I might raise the wrong hand by mistake, or try unsuccessfully to raise my hand against resistance). I do not have to do anything to bring<sup>it</sup>/about that my hand goes up. So the satisfaction of the desire that the desire to raise my hand should move me to act is not a means to the satisfaction of the desire to raise my hand. It is in normal cases a way (in fact, the way) of satisfying the desire to raise my hand. Furthermore, the satisfaction of the desire to raise my hand is the way of satisfying the desire that the desire to raise my hand should move me to act.

There might be some doubt about this analysis, in that it seems to subordinate the second order volitions to the first order desire: the second order volition only has motivational efficacy because it happens to reinforce the first order desire. To avoid giving this impression I want to consider a somewhat more artificial model for the efficacy of second order volitions which shows how we can, in certain cases, think of the second order volition as playing a more prominent role. Miss Anscombe (89) writes; "It is a mistake to think that one cannot choose what desire is going to be motivated by - Plato, saying to a slave, 'I should beat you if I were not angry' would be an example." This point has been criticised by Kenny (90), on the grounds that, although one might be able to choose between different acts on the basis of different desires, once one has decided what one is going to do,

one cannot then raise the further question of which desire is going to move one to do it, except possibly by deliberating whether to perform some further action which will set the first action in a different light. (91) In what sense can Miss Anscombe's example be described as a case of someone's choosing his own motivation? Presumably Plato has three relevant psychological attitudes in this case - his anger at the slave, something like a moral condemnation of the slave and a desire, again probably also a moral one, not to be motivated by sheer anger. His primitive anger at the slave and his moral condemnation of the slave both lead to a desire to beat the slave. The principles governing the individuation of desires are a matter for separate discussion; but I take it we would not speak of his having two desires to beat the slave; rather there are two sources of his desire to beat the slave. Plato feels that if he were to act on this desire he would, at least in part, be acting out of anger. He is able, we are to suppose, to refrain from beating the slave and his reason for thus refraining would be his desire not to act out of anger.

The general structure of this example would seem to be this: someone refrains from acting on a certain desire because of a desire not to be motivated by such a desire. Frankfurt uses the term "second order volition" to characterise desires that a certain desire should move one to act. I shall use the term "second order counter-volition" to characterise desires that a certain desire should not move one to act. This might help us with our problem: can one act on a desire and one's acting on this desire be on the basis of the desire to act on this desire? We had difficulty in producing an account of how this might be possible, which assigned anything more than a subsidiary supporting role to the second order volition. By contrast Miss Anscombe's example does seem to be a clear-cut case of a second order counter-volition playing a dominant role in an action or rather in someone's refraining from an action. For it is the only desire on one side of the conflict and yet it is the one that wins out.

Suppose someone has two conflicting desires for X and for Y and a second order volition that the desire for X should effectively move him to act. He believes he can satisfy either but not both of them; it is in this that the conflict between his first order desires consists. Now his second order volition that the desire for X should move him to act would in this situation quite naturally lead to a second order counter-volition that the desire for Y should not move him to act. As we have seen, it is quite intelligible that

someone should refrain from acting on the desire for Y because he desires that he should not be moved to act by the desire for Y. (When I describe the possibility as intelligible, I do not mean that I have provided an analysis of it, but simply that the possibility seems to be one that actually occurs.) So the person can secure the conformity of his will to his second order volition as follows: his second order volition to act on the desire for X leads to a second order counter-volition not to act on the desire for Y, which leads him to refrain from acting on the desire for Y. This, as it were, clears the way for his acting on the desire for X, and it is not merely a happy chance that his acting on the desire for X conforms to his second order volition.)

This model, I have said, is artificial. For one thing, it is not obvious that the process must go via the formation of a second order counter-volition. And for another, it seems to postulate two separate acts: one of refraining from acting on the desire for Y and one of acting on the desire for X. Such cases may occur but they seem somewhat *récherché*. The model does however serve its purpose: to show that if we think of the second order volition as reinforcing the first order desire, this does not commit us to regarding it as merely another desire along with the rest with no part to play of its own. Miss Anscombe's example shows that a second order counter-volition can be effective on its own and as a second order volition is such that by practical reasoning it could give rise to a second order counter-volition, it would seem that it could have the motivational efficacy possessed by the latter. It could be the decisive factor in preventing the efficacy of those first order desires which conflict with the favoured first order desire. It seems that a reasonable case can be made out for the efficacy of second order volitions. Finally, it should be noted that a second order volition that a first order desire should be the only desire that moves one to act in a certain situation could not be efficacious as, if it were, it would itself be contributing to the action, thereby defeating itself.

We have distinguished two types of second order desire, those where one desires to have a desire which one does not already have and second order volitions, where one desires that one of one's desires should move one to act. There is considerable doubt about the efficacy and to some extent the intelligibility of the former, whereas the latter seem conceptually acceptable. Earlier it was suggested that there would be considerable difficulty in justifying one's acceptance of a putative case of the efficacy of a

a second order desire of the first kind; because the demand for some evidence for the appearance of the desire which was wanted, if met, would then lead to a suspicion that one had misdescribed an example of a second order volition's proving effective. There is no particular temptation, at least in our analysis, to think of the efficacy of a second order volition as involving a mental act (except possibly if one regards refraining as a mental act, a view against which I shall argue shortly). As I said earlier the object of this chapter is to argue that a somewhat different approach is needed when we come to mental acts, from that which sufficed for physical actions. We shall miss much that is significant about mental action unless we approach it <sup>expressly</sup> ~~especially~~ from the standpoint of what we can and cannot do, or rather of what it would be intelligible for us to be able to do. Let us now return to the problem of desire and belief at will, and in doing so generalise it to one involving other mental states.

We saw earlier that if one were able to desire something because one wanted to desire it such a desire would be adventitious. It could not be justified by reference to one's other desires, fears or values in the normal way. The object of this desire would not stand on such a relation ~~as~~ that of means to end to the objects of any of one's other desires, or, if so, only by accident. If one were to attempt to justify the desire one would have to say not what was good about the object of the desire but what was good about having the desire. It is not clear that one can immediately conclude that the desire is somehow spurious because the justification for it is of the wrong type. It might be ~~that~~ we have simply unearthed a new way of justifying a desire, a new kind of practical reasoning. More importantly, not all the desires we have are one's that spring from further considerations or are justified by reference to them. Might it be that the desire ~~is~~ one for which one has no further justification rather than a justification of the wrong kind?

Miss Anscombe writes (92), "But is not anything wantable, or at least/<sup>any</sup> perhaps attainable thing?" It will be instructive to anyone who thinks this to approach someone and say: "I want a saucer of mud" or "I want a twig of mountain ash". He is likely to be asked what for; to which let him reply that he does not want it for anything, he just wants it. It is likely that the other will then perceive that a philosophical example is all that is in question, and will pursue the matter no further; but supposing that he did not realise this, and yet did not dismiss our man as a dull, babbling

loon, would he not try to find out in what aspect the object desired is desirable? Does it serve as a symbol? Is there something delightful about it? Does the man want to have something to call his own, and no more? Now if the reply is: "Philosophers have taught that anything can be an object of desire; so there can be no need to characterise these objects as somehow desirable; it merely so happens that I want them", then this is fair nonsense.

"But cannot a man try to get anything gettable? He can certainly go after objects that he sees, fetch them and keep them near him; perhaps he then vigourously protects them from removal. But then, this is already beginning to make sense: these are his possessions, he wanted to own them; he may be idiotic, but his 'wanting' is recognizable as such. So he can say perhaps; "I want a saucer of mud". Now saying "I want" is often a way to be given something; as when out of the blue someone says "I want a pin" and denies wanting it for anything, let us suppose we give it him and see what he does with it. He takes it, let us say, he smiles and says; "Thank you. My want is gratified." - but what does he do with the pin? If he puts it down and forgets it, in what sense was it true to say that he wanted a pin? He used these words, the effect of which was that he was given one: but what reason have we to say he wanted a pin rather than to see if we would take the trouble to give it to him?

"It is not a mere matter of what is usual in the way of wants and what is not. It is not at all clear what it meant to say: this man simply wanted a pin. Of course, if he is careful always to carry the pin in his hand thereafter, or at least for a time, we may perhaps say: it seems he really wanted that pin. Then perhaps, the answer to "What do you want it for?" may be "To carry it about with me" as a man may want a stick. But here again there is further characterization: "I don't feel comfortable without it; it is pleasant to have one" and so on. To say "I merely want this" without any characterization is to deprive the word of sense; if he insists on 'having' the thing, we want to know what 'having' amounts to."

Miss Anscombe appears to feel a certain uneasiness about these cases of unintelligible desires. On the one hand, it seems that the criteria for someone's having some such desire as that his best friend's brother-in-law should have a bank-account in Birmingham might be satisfied as <sup>w</sup>ell as could be, and yet on the other it could be that the person claims to have no further reason

for desiring this - the situation specified simply seems desirable to him on its own account. The desire seems unintelligible; and yet it is difficult to reject it, to deny it the status of a desire. Surely it can hardly be a conceptual truth that in such a situation the man must be lying, or else have unconscious reasons for his desire. Admittedly if someone simply says that he wants a saucer of mud, but will not say what counts as getting it or what he proposes to do with it when he gets it, then we might have conclusive grounds for regarding the object of his desire as inadequately specified, and therefore not feel bound to accept any unpalatable conclusion about his having some particular unintelligible desire. But suppose he answers our question as to what he wants a saucer of mud for and what counts as getting it, and yet seems no more comprehensible than before - for example, "I want it to balance on my head."

There does not appear to be much benefit to be gained from dismissing a desire as idiotic or unintelligible, if it satisfies all the criteria (action, avowals etc) for being a desire and if its object is adequately specified. Such a dismissal cannot justifiably carry the implication that one is not dealing with a genuine desire. It seems that desire at will is likely to involve unintelligible desires of this sort, particularly as regards the second order desire. To say that a desire is unintelligible is not perhaps to say that it is inexplicable. Indeed the fact that some unintelligible desires might be explicable is a strong argument for accepting them as genuine desires: imagine someone being given a post-hypnotic suggestion that he should want his best friend's brother-in-law to have a bank-account in Birmingham. If desire at will is possible, then the desire which thus came into existence would be explicable though probably not intelligible.

I think it is fairly clear that there is not likely to be an argument against the possibility of desire at will which has quite the cogency of those against belief at will. Beliefs are something for which certain types of reason are appropriate: the wrong sort of reason or no reason in some cases will simply prevent us from regarding it as belief. I do not wish to imply that for all beliefs one must have a reason; but those justified beliefs for which one has no reasons, like beliefs as to the position of one's limbs when one cannot see them, would become problematic cases of belief if reasons were offered in the form of desires to believe. Desire at will, on the other hand, seems pretty unintelligible, but this is not a kind of unintelligibility that clearly implies inconceivability.



These considerations as to the type of justification appropriate to mental states serve as a basis for a treatment of states other than desire and belief. If one justifies one's anger or resentment at someone one will state one's beliefs as to what he has done and one's reasons for thinking what he has done to be undesirable (mentioning such things as one's overwrought state are more in the nature of excuses than justifications). This immediately sets limits to the extent to which anger or resentment can be subject to the will. One cannot, for example, decide to be angry with Smith for stealing one's property unless one believes he has stolen one's property, and this is not subject to the will. But given that one believes that Smith has stolen one's property and one does not want one's property stealing, can one then be angry or not angry with him according to whether one wants to? It seems that it might just happen that one is not angry even though one has good cause but could this be because one does not want to be angry? And furthermore, it might be <sup>more</sup> that one does not act on one's anger because other considerations prove stronger.

I shall not attempt a full answer to this question, primarily because it would involve an analysis of anger to determine to what extent anger is an isolable feeling, certain desires and actions being simply the results of this feeling. Some analyses are obviously wrong, for example, any that treat the connection between anger and the desires to which it leads or which manifest it as just contingent. But nothing follows immediately from this as to whether there is a specific feeling of anger. According to the position one takes on this issue one will give differing answers to the question as to whether anger is subject to the will. If one regards anger as involving behaviour, rather than merely that which is liable to be followed by a certain kind of behaviour, then to what extent it is subject to the will: for one may or may not behave in certain ways when one believes someone has wronged one, and this will in part depend on desires other than those arising from anger. If one analyses anger in terms of certain desires but not the behaviour to which these lead, then anger at will ought to present the same sort of difficulties as desire at will. If one regards anger as a feeling then one will face the problem: perhaps some feelings (and one needs to remember the range of this word) might conceivably be subject to the will but anger is a feeling because of something. It is unclear how this could be brought about by a desire to be angry. The difficulty is evident if the "because of" is analysed as a causal relation, a relation which is discovered rather than created. This seems to present a very great obstacle to the idea

of anger at will, on this last interpretation (93).

Other emotions and affective attitudes will present the same problems. An immediate limitation will come from the fact that the justification of one's emotions involves the presentation of one's beliefs and desires. If one then modifies the question to that of whether, given that one has beliefs and desires which are of the right sort to justify a certain emotion, one can choose to have the emotion according to whether one wants to, one will get varying answers depending upon the position one takes as to what the emotion itself consists in. A large number of mental states apparently are not easily regarded as subject to the will. Is there any pattern in this? The mental states we have considered have all been such as to require certain specific modes of justification: and to regard someone as being in them because he wants to be tends to undermine this justification to varying degrees. Thus if a mental state is a cognitive, or a conative one (desires, fears, values: roughly, the category which Kenny calls Volitions), or if the state presupposes for its existence the existence of such states and must exist because of them (in the way affective states and emotions do) then this at once imposes severe, perhaps prohibitive, constraints on the possibility of someone being in such states simply because he wants to be.

So if we want to find examples of mental actions we had better look elsewhere. Those mental states which are neutral with respect both to fact and value furnish promising examples. Simply thinking about something, entertaining a proposition or having a mental image are not cases of having beliefs or desires, nor do they presuppose them in the way affective states do. Furthermore, if we ask someone for his reasons or justification for thinking about something, entertaining a proposition or having a mental image then, if one gets an answer at all, it is likely to be some merit he sees in his having such thoughts or the consequences of such thoughts: "I am trying to solve this problem", "I am trying to decide whether the proposition is true", or "I am trying to remember what kind of shoes Smith wears". One might, of course, get no answer in terms of his reasons, the thought might have just come to him, but if he does attempt a justification it will not be an argument for the truth or falsity or desirability or undesirability of the state of affairs envisaged but for the desirability of his contemplating such a state of affairs.

This is the right kind of justification one needs if the mental states are to be regarded as the upshot of mental actions.

Consider the case of someone who is trying to solve a problem of mental arithmetic. In the face of various distractions he keeps his mind on the problem, and redirects his attention to it, whenever it wanders. His thinking about the problem he would justify by maintaining that only then would he be likely to solve it. Such a case may well be paradigmatic of mental action. If there is one type of mental state that is clearly subject to the will it would seem to be that of attention. Suppose the man suddenly sees the solution, or understands how the problem should be tackled. Has he performed a mental act? The answer has to be in the negative, I think, for it makes no sense to qualify the verb "see" and "understand" in this case by adverbs like "intentionally" or "unintentionally" (although one could describe the events as intended). Nor can one ask him for his reasons for understanding (although there might be reasons why he understood or why what he understood was intelligible). It might well be that his understanding was the result of his previously trying to understand; and here we do seem to have a case of action. One could ask someone for his reasons for trying to understand something. Admittedly it sounds odd to speak of intentionally or unintentionally trying to understand something, but this seems to be a feature common to all cases of trying. If one is trying to do something, one's action must be intentional under that description.

Trying to understand something, as far as I can see, involves the direction of one's attention. A similar point could be made with regard to remembering and trying to remember. Or reaching a decision and trying to reach a decision (where the decision could either be as to a matter of fact or a matter of policy). Could it be that all mental acts involve the direction of one's attention? We have seen that the mental states to which we seemed bound to turn if we were to discover genuine cases of mental action were non-committal as to fact or value, and this left it open that <sup>one</sup> ~~we~~ should be able to justify one's being in such states by saying what was desirable about being in them. Thus if a desire to be in such a state causes one to be in such a state one is not faced with any immediate difficulties <sup>for</sup> ~~about~~ this being a case of action. And it seems that all such cases will involve the direction of one's attention. One will no doubt cavil at some examples of this: if someone desires to think of a house and this causes a thought of a house, then there seems something odd about this. Is he not indesigning to think about a house, already thinking about what he wants to think about? But one can avoid

this problem in particular cases by formulating his desires as "a desire to keep thinking about a house", "a desire to have a mental image of a house" etc. and such desires seem to be desires concerning the direction of his attention and their efficacy consists in his attention becoming directed as desired because of them.

Such considerations do not of course amount to a proof. To provide one would require nothing short of a full treatment of all mental states and events. Important omissions in our treatment include: the efficacy of desires not to have a certain desire, of desires to experience certain sensations and of desires to experience feelings which are not localisable sensations. Finally we have not dealt with the vast question of one's attitude to one's mental states, in particular to one's emotions; whether this is one of whole-hearted participation, grudging acquiescence or resolute opposition. But enough has been said to indicate that one must not expect mental acts to be particularly thick on the ground: it might well be that all mental acts are a matter of one's directing one's attention. Whether this is, in itself, a severe restriction is not obvious; perhaps the restriction comes more from the fact that so many mental states require a justification which is not easily reconciled with a justification in terms of the desirability of one's being in such states. Or perhaps, as the particular course taken by our argument has suggested, these two points are connected. It is noteworthy that attention is not always, perhaps not even usually, directed purely to one's thoughts, to one's inner life but to objects and events in the external world. If the only forms of mental action involve the direction of one's attention and if there is some conceptual primacy of attention to the "outer" as against attention to the "inner", then it may be that action itself is primarily action directed on to the external world (not necessarily action in the external world ~~as~~ attending to a physical object is not to affect or modify it).

Let us end this discussion by considering two further putative cases of mental action which would not have fitted neatly into the above discussion. A treatment of the first (the case of refraining) is required because the concept has already been employed in our analysis, and something must be said of the second (a certain possible kind of trying) because it threatens to make difficulties for our analysis of action generally.

Someone is tempted to steal various goods when in a supermarket. He also has reasons for not doing so and because of these latter, he refrains from stealing the goods. There seems little doubt that intentionally refraining from doing something is an act of some sort. It is intentional, one has reasons for it and one speaks of acts of refraining and forbearance. But is it a mental act? The obvious reason for regarding it as a mental act is that otherwise, unless we create a new category of actions which are neither mental nor physical we must regard it as a physical act and yet there seems no relevant bodily movement which occurs, no physical <sup>event</sup> ~~act~~ as part of the allegedly physical act. However, there does not seem to be any mental event which fills the bill either: if someone ~~fe~~refrains from doing something there is neither a mental nor a physical event which can be isolated as the event-component of the act, in the way that one's arm going up can be extracted from one's raising one's arm. The man in my example desires to refrain from stealing some goods; for this reason he refrains from stealing them. It is hard to see what the mental event (or state) could be, which if it occurred (or arose) because of the desire, would give us a case of someone's refraining from some physical action. In fact, it is clear that if refraining from some physical action were a mental act of this sort, then one could refrain from doing something, and yet do it: an absurd conclusion.

So it looks as though we are going to have to regard refraining from a physical action as a species of physical action, unless we accept it as *sui generis*. Let us therefore endeavour to remove any suggestion of paradox that might seem to attach to this conclusion. We clearly cannot analyse refraining as a desire causing a bodily movement when there is no relevant bodily movement to be found. It will have to be thought of as a desire causing a bodily state, but not the coming into existence of a bodily state. Thus if I refrain from raising my arm, some desire of mine causes my arm to remain motionless. Is this a coherent notion? Can something be caused to stay as it is? Intuitively there seems nothing wrong with talking about the cause of the bridge's failure to collapse, or of a car's remaining motionless when one attempts to start it. Is there always, in such cases, ~~some~~ event which one believes would have caused some change, if it were not for other factors which one judges responsible for things remaining the same? No, because one could ask for the cause of something's remaining suspended in mid-air, without implying the occurrence of some event which ought to have precipitated its

falling. But perhaps one does imply that there is some state of affairs which would lead one to expect it to fall.

How does this affect our assessment of the statement that someone refrained from something? More particular if we say that someone refrained from doing something for a certain reason, do we wish to imply that, had he not had that reason, he would not have done it? Suppose I say, "He refrained from signalling (by raising his arm) because he believed it would have hurt his arm". Apart from cases where the person has other reasons for not raising his arm, in which event my remark would be, as it stands, misleading, there is an implication that if he had not believed signalling would have hurt his arm, he would have signalled. I would suggest that this is a general and that someone's reason for refraining from doing something is a reason for refraining from something he would otherwise have done. If, out of the blue, I am asked why I refrained from flying to Bangkok yesterday, it would normally be misleading for me even to attempt to answer the question in that form unless I had actually been contemplating going to Bangkok or had had some reason to do so. One cannot regard just anything a person did not do as involving an act of refraining on his part; it must be something he had a reason for doing. (Probably, it is also the case that when someone refrains from doing something, he must refrain for a reason. Forgetting to do something one had a reason for doing is not refraining.) There are, of course, cases where someone acquires a reason for not doing what he felt no inclination to do anyway. But I do not think we would even speak of his refraining from doing it under these circumstances.

The kind of refraining we have been considering, the one that is problematic for us, is that where there is no overt bodily movement. It is not of course the only kind. If someone is deliberating whether to do A or B, where these both involve physical action and where he believes he cannot do both, then, if in the end he does A, he by the same token, refrains from doing B. But this does not involve us in the difficulty of there being no bodily movement to which we can refer. This is not to say that such a case poses no analytical problems, but at least there is no temptation to speak of a mental act here. It should also be noted that we are giving an account of refraining, not of omitting.

Omitting to do A is something one can do without having a reason for it, nor need there be anything which one does for a reason, in doing which one ipso facto omits to do A. But, again, it does not seem likely that there will be any temptation to regard those acts of omission which are not acts of refraining as mental acts.

The second case I want to consider is the following:  
someone decides to give a signal by raising his arm. Unknown to him his arm has become paralysed. He raises his arm, so he thinks, but then discovers by observation that his arm has not moved. What, if anything, did the man do. (95) Did he do something which fell short of raising his arm or did he merely think he raised his arm? Supposing he did something which fell short of raising his arm, let us say he X - ed. Could not X - ing be susceptible to paralysis, and could one not construct a case of an unwitting paralytic similar to the above? If this is conceivable then someone might think he has raised his arm when he has not, but more over when he has not even X - ed. Perhaps he has done nothing at all or perhaps he has Y - ed, something which falls short both of X - ing and of raising his arm. If the latter, then can we not construct another example in which Y - ing is susceptible to paralysis? There is nothing vicious about this regress. Anyone who takes the view that our original unwitting paralytic does something, is not thereby forced to say this about the further cases we construct, nor is he even forced to admit them as possibilities: he could deny that one could lose the ability to X, or that one could fail to X without knowing it. What the speculation does show is that the notion of the man just being wrong to think he has raised his arm, without there necessarily being something he did, is a somewhat simpler explanation of the situation. What is more, one will have to say this about some of the cases we have constructed if one is to block the regress. If one says that our unwitting paralytic in fact Xs, then faced with a case in which he sincerely claims to have X - ed but in fact has not, one will have to say either that he merely believes he has X - ed, or he has Y - ed which fell short of X - ing. And at some point one must terminate the regress by claiming that he is simply wrong. Unless, that is, one can unearth an incorrigible element in a man's knowledge of his own actions.

Let us investigate this last suggestion. Perhaps if someone thinks he is doing or has done something then (so long as in the

latter case we do not have an example of failure of memory) there is something he does or has done. So from the statement that someone believes he is raising his arm, we can infer that he is doing something, even if it is not raising his arm. Perhaps indeed we can infer something more definite: that, for example, he is performing or has performed some kind of act of will, or a kind of trying. (This more specifically<sup>view</sup> would be natural if one assumed that what the unwitting paralytic did in our example was also done by a normal man raising his arm, as an effect of which his arm rose). Does it follow from the fact that someone thinks he is doing something, that there is something, perhaps a different something, that he is doing? (96)

It is not clear why this implication should be thought to hold. There seems no reason why someone should not be completely mistaken as to whether he is performing an act. It is conceivable that a neurophysiologist should find a way of stimulating a person's <sup>brain</sup> hair so that he wrongly thought that he was moving some part of his body (and perhaps produced a rationalisation for this). Must we then assume that the neurophysiologist has found a way of producing certain acts? The view under discussion represents action as being in a way like pain. We allow some latitude for a person to be wrong about the location of his pain, say, but not as to whether he is in pain (given that he knows how to use the word). Similarly, it is suggested, one can be wrong about whether one is performing a particular act, but not about whether one is acting.

Now it may well be true that we have no use for the notion of an illusory pain. But the same cannot be said of the notion of an illusory act. If someone thinks he is doing something A, we certainly recognise that he might not be doing A. Thus the situations are not strictly analogous. If we assume that there must be something he is doing, which falls short of doing A, the grounds for this will have to be different from those for the seemingly similar thesis about pain. Consider the following point<sup>s</sup>:-

(a) If someone is fully aware that his arm is paralysed, it is not clear that he can even try to raise it. Human beings are not capable of moving their hair without doing anything to bring about the motion of their hair, and it does not seem that they can try to do so either.



(b) It seems to be contingent that sometimes when an unwitting paralytic is ordered to raise his arm he believes he is complying, when in fact his arm remains motionless at his side. If human beings had no experience of paralysis, that is the idea of their losing their power to perform the bodily movements they could perform was merely a logical possibility to them, what basis could they have for predicting that paralysis would involve the phenomenon which interests us?

(c) If when an unwitting paralytic falsely thinks he is raising his arm, he is actually doing something, it looks as though his action is going to be unanalysable. What event does his desire to raise his arm or that his arm should go up, cause? There is no event to which we can point. We are admitting a type of action in which there is no isolatable event-component. The difficulty arises whether we think of what the unwitting paralytic does as a mental act or (implausibly) as a physical action.

~~Not~~ Now (c) is clearly a difficulty for us. If we are to defend a causal theory of action along the lines already laid down, we cannot afford to admit actions with no event component. If we are to analyse action as (roughly) a desire causing an event, any action which apparently involves no particular event which could be considered as that caused by the desire (or at least any event the desire does cause must await discovery by physiologists) is going to prove an embarrassment. Suppose it is said that the unwitting paralytic tries to raise his arm. Then it seems that this "trying to raise his arm" is not easily broken up into desire plus caused event. Our example does involve the occurrence of a specific event: the man's coming to believe he is raising his arm. But it would be odd to use this event as the one needed, on our analysis, to be the result of the desire. For one thing, it is not suitably related to the object of the desire. (What sort of practical syllogism would exhibit a suitable connection?) The same difficulties arise if we try to analyse the action as the desire causing a state (which, anyway, implies the coming into existence of the state) for there is no relevant state to which we can refer. These considerations are not really strong arguments against the interpretation of the example as one involving an action; they rather constitute an admission of the fact that such an interpretation does not fit in well with our analysis. So let us pass on to a consideration of (b) and (c). (a) ?

Danto gives an example of someone who takes a drug which gradually paralyses his arm. (97) It gets more and more difficult to move his arm: he has to try harder and harder to move his arm. But eventually he cannot even try to move it. Presumably, if he did not know he had taken the drug, and did not know that his arm was gradually becoming paralysed because he had not tried to move it in the early stages, he might eventually have the experience of wrongly thinking he was moving his arm. If this experience involves his actually doing something, it is hard to see why he should only do this when he is unaware of his paralysis. (98) If it merely involves his having a false belief, it is not difficult to see why this should be so. For to suppose otherwise would be to countenance the man's having contrary beliefs. Consider<sup>action</sup><sub>a</sub> (b) thus supports the view that the person does not perform an action.

So, I think, does consideration (b). It is conceivable that there should be a race of people who never have the experience in question even though they are sometimes victims of paralysis. There would be no temptation to say that they performed any mental acts of trying or willing which somehow failed to issue in a bodily movement, or any occasions on which they were paralysed. Why should we say that human beings who do have this experience are performing an additional mental act, rather than that they are merely susceptible to a certain kind of false belief to which the people of our hypothetical race are immune? Suppose that, owing to some mutation, there began to appear among this race the odd case of the experience we are discussing. Would there be any reason to suppose that the mutants were not merely susceptible to a certain kind of false belief, which their predecessors had escaped, but that they were performing a new kind of act of which the race had previously been aware? The members of the race themselves might argue: "Mistakes of this degree of severity about one's own actions are unheard of. Surely these people must be right about something. They must have performed some action, possibly a mental one." But even this would be a suggestion from precedent; it is far from a logical truth. We, who are not members of this race, do not have even precedent to go on. We know that people are sometimes mistaken about whether they are performing a certain action even when there is a basic action. The contingency of the fact that the experience in which we are interested occurs seems to point to its not showing some fundamental truth about the nature of action, and the most economical account, one that does not introduce

a special type of mental act, is probably the best. (99)

I have not in this chapter stated an explicit analysis of mental action. But enough has been said to show that such an analysis (of a genuine mental act) would parallel that of physical action. Firstly, if a desire is to explain a mental action as a reason for that action it must match that action in the practical syllogistic way discussed in chapters 2 and 3. Secondly, there will arise once again the Davidsonian demand for an explication of the "because" in such statements as "He concentrated on the problem because he wanted to understand it". The trouble is that an analysis based on these considerations<sup>alone</sup> allows all conceivable or seemingly conceivable cases of a desire to be in a certain mental state (or a desire of some other kind, together with a belief that being in a certain mental state would be a means to satisfying it) causing that mental state to count as possible actions. I have tried to show that the nature of many mental states puts further very severe restrictions on what could count as a mental act.

Chapter 8

The Explanation, Understanding and Criticism of Action

Suppose someone performs an action A and we ask him why he did it. He replies, "Because I wanted X". We, however, reply, "But you also wanted Y and could have had that instead. Admittedly, you knew you could not have both X and Y. But why did you choose to act on the desire for X rather than the desire for Y? You haven't fully explained your action until you tell us why you made one choice rather than the other." Suppose the person continues: "Well, I wanted Z as well and performing Action A is a means to getting Z." (100) But we are still not satisfied: "Why did you choose to act on the desires for X and Z rather than on the desire for Y?" Clearly, if we continue in this vein much longer we will show ourselves unwilling to accept any explanation of the action which merely mentions desires which the performance of A would contribute to satisfying. And, the same treatment would be meted out to fears, moral values, beliefs about one's future desires and beliefs about the desires of others if these were adduced in explanation.

Must the man give up in exasperation at this point or is there anything else he can contribute in explanation? He might, I suppose, mention second order volitions such as the desire that the desire for X should effectively move him to act. But assuming that such entities can have motivational efficacy (if they cannot then the man would be wrong to offer them as an explanation of his action) we can subject them to the same treatment as before. A desire that the desire for X should effectively move him to act would be in conflict with the first order desire for Y in this situation and we can ask him why he acted on this second order volition together with the other desires rather than the desire for Y. After all, presumably a second order volition does not have to be effective; so one could press one's series of questions by asking why it moved the man.

Thus the introduction of second order volitions does not effectively block the sequence of questions. Such an interrogation may be perverse but my point is that it is not more perverse to reject the explanations as insufficient when second order volitions are invoked than when the explanation consists of a set of first order desires. The person might offer another kind of explanation in terms of the relative probability of his being able to satisfy

his various desires. Thus he might say that he did not act on his desire for Y because he was not certain that whatever it was that he was contemplating as a means to getting Y could in fact have the desired effect; whereas he had much greater conviction that performing A would lead to his getting X. Let us assume, for the sake of simplicity that the man had no beliefs concerning the differential likelihood of his desires being satisfied. One could continue the questioning even where this assumption is not true e.g. "Why did you act on your desire for X and belief that X was easily attainable, rather than your desire for Y which you believed was less certain to be attained?" For people do sometimes choose to act on the desire that has less chance of satisfaction, rather than on some other, the fulfilment of which is more likely.

An objection that may be made at this point runs as follows:- when a person has done something wrong it is common to demand, "Why did you do it? Why were you moved by such-and-such a consideration rather than by your moral convictions?" But when someone has done something he believes right (at least if we also believe it right) we do not enquire, "Why did you do it? Why did you act on your moral convictions rather than on your baser desires?" But the objection is not well taken as the reason why we rarely make such an enquiry, ~~It~~ seems to be that we regard a person's acting on his moral convictions as self-evidently justifiable, (although we do not necessarily regard a person's particular moral convictions as justifiable). Our obsessive interrogation need not be a demand for justification, but simply a demand for a fuller account of the reasons why the man acted as he did. People do not always act on their moral convictions rather than on conflicting considerations and we may want to know why the man did so in this case. Further reasons might indeed tend to weaken the claim that the man was acting or acting solely on his moral convictions; but the fact remains that we might simply be puzzled by the fact that the man acted as he did.

This suggests the first lesson to be drawn from our interrogation: although there must come a point where the man can give no further reasons, this point need not come at any particular stage. Or we might say that there is no logical limit to the number of reasons someone might have had for a particular action. The second lesson concerns the degree to which an action can be explained by giving the reasons for which it was performed. For it seems that, in some cases at least, we might know all the reasons a person

had for a certain action and yet still be puzzled, and this not because we are puzzled by the reasons themselves.

An action has been completely explained in terms of the agent's reasons when the reasons for which he performed it have been fully enumerated. If a certain desire is mentioned, then one may require or be able to give an explanation of why the person has this desire, and if this explanation is in terms of further desires, to the satisfaction of which the object of the desire in question is, for example, a means, then one might include this as well. Perhaps such reasons are also, in principle, capable of complete enumeration. But one will not mention anything, which although it might in some way explain his action, could not be construed as his reason for acting. For instance, one would not include a biological explanation of why it is that human beings are subject to hunger. Suppose one has such a list for any particular action. If the agent had reasons for not performing the action in question one can also enumerate these. If a person has reasons both for and against a certain action, this is consistent both with his performing that action and his not performing it. An explanation in terms of an agent's reasons can be complete, but we are dealing with a completeness of a special sort; merely from knowing that he had those reasons, one could not have predicted his action. His having those reasons is compatible with his not performing the action. If the list of reasons is supplemented by a complete list of his countervailing reasons, the position remains the same; indeed it is underscored. And yet we feel that it is not inappropriate to describe the complete list of the agent's reasons for performing the action as a complete explanation of his action. Or, to put the point another way, although (for most actions at any rate) when one says someone performed an action one has to be able to answer or to assume answerable the question: "Why did he do it?", one does not have to answer ~~or~~ at any level such questions as, "Why did he do it, rather than something else?", even when one knows that he had reasons for doing something else. Nor is it obvious that one must assume such questions answerable. This seems a radically different kind of complete explanation from a complete determinate<sup>istic</sup> explanation.

I have refrained from putting the point in terms of the intelligibility of actions. One might indeed be nonplussed by the fact that someone acted on certain desires rather than others, but this need not always be the case, even where one feels that one

would not oneself have made the same choice. So it would be misleading to say that we have discovered a limit to the intelligibility of actions. This is particularly so as it can be thought of as stemming ultimately from the truism that a person cannot have performed an action for more reasons than he did. And certainly, one should avoid such grandiose formulations as "the inconceivability of human action according to the Principle of Sufficient Reason".

Two points need mentioning before we proceed further. We do provide explanations of why someone acted on the desire he did by saying how that desire came to be so strong. We sometimes offer explanations of how certain considerations came to weigh so heavily with someone. Such explanations might be couched in psychological or sociological terms. And it might be an aim of psychology to provide explanations of this sort for all actions. Similarly, any determinate<sup>istic</sup> explanation of human behaviour must aim at explaining behaviour in a way that does show why a certain action (or perhaps a certain bodily movement) occurred rather than some other, although it is not obvious that it must do this by involving the concepts of desire and of the strength of the desire. This must be admitted; but such explanations do not explain by giving the agent's reason for his action. They may take as their starting point the agent's reasons for and against an action,<sup>and</sup> explain a choice by considerations of the relative strengths of the desires, fears, moral convictions etc. but this is not giving reasons alone.

The second point is that we often attempt to explain actions by speaking of preferences or by saying that the agent felt that a certain course of actions was best. This is particularly common when one is explaining one's own actions. Now, again, this is not to offer reasons; though it might be to say that certain reasons, in some sense, outweighed others. But furthermore, it is apt to degenerate into triviality. It is well-known that there is a reading of "Someone acted on a certain desire because that desire was the strongest" that is blatantly non-explanatory: when one takes it as the criterion of a desire's being the strongest that it should be the one that leads to action. There are similar interpretations of "He acted on the desire for X rather than the desire for Y because he preferred to" and "He acted on the desire for X rather than the desire for Y because he thought it was for the best." Such locutions need not be completely vacuous. They may tell us that a man's actions accorded with his previous judgements as to the strength of his desires, or as to his preferences or as to what he thought best. But this is hardly a contribution to the further explanation of his action.

If anything, it suggests the possibility of further questions to be asked: for had the man's action not been in accordance with his previous judgements, problems would arise as to why this was so. It is probably not contingent that a man's judgement as to his preferences before he acts, or when he has no opportunity to act, generally accord with how he does act when he is presented with a choice between courses of action there and then. In a case where there is no such accord, this would itself present a problem. (101)

There is a class of explanations of actions which are similar to those discussed above but which seem more specific as to the way in which the action in question was considered more desirable. For example, the man might say, "I thought I would enjoy X more". Now, if the reason he had for and against the action involved considerations other than enjoyment, such an explanation would be just so much grist to the mill of our interrogation, for we could ask, "Why did you choose to act on considerations of enjoyment, rather than on the others?" The action would clearly have been explained only in the sense discussed earlier, in which a complete statement of the reasons the man had for acting as he did would be compatible with his having acted otherwise.

On the other hand, if the considerations which weighed with the man for and against the action involved no considerations other than enjoyment, one might feel that the answer "I thought I would enjoy X more" closes the gap. It would at least be odd in the circumstances to ask, "why did you act so as to secure that which you thought would give greater enjoyment rather than that which you thought would give less?" But here it seems that the man's choice is already being made under a particular heading (102): which action will give greater enjoyment? If so, it does not seem particularly different from deliberating or choosing how to act where one has already decided upon one's objective, but still remains to decide upon the best means of obtaining it. Such a decision might turn upon one's belief as to a particular matter of fact, and here it seems that the matter of fact concerns one's relative enjoyment of certain situations. To say that this is not really practical deliberation as the man has already determined upon a particular policy and is concerned only with the factual consideration as to the most certain or the most efficient mean to implement it, would be too strong; for he is deciding what to do. Nevertheless, this kind of case is not of the type which interests us, that of



choosing between rival goals. If the man is already choosing between actions according to which he believes will give greater enjoyment, it would seem that the question as to which of his desires, fears, moral values etc. will govern his choice has already been decided (he may not have decided upon them, as he may never have raised the question as to whether to act on considerations other than those of enjoyment).

The cases which we want to consider are those in which we are dealing with a persons action where the reasons he had for performing the action and the reasons he had for not doing so cannot be brought together under any non-trivial heading. His choice or decision (if his action was preceeded by a decision) cannot be interpreted any more specifically than as choosing or deciding upon the course of action that he preferred, whose object he wanted most or he thought was for the best. Suppose the person who did A did it because he thought it would lead to X which he expected to enjoy, and because he thought it would avert W which he feared, but that he also had reasons for not doing A which included the fact that he believed that if he did not do A, he could then do B which he felt morally obliged to do. In this situation there is no non-trivial heading under which his choice occurs.

We have said that from a full statement of his reasons for doing A, one could not have predicted with certainty that he would do A. His having those reasons is compatible with his not doing A. This is true in that there is no entailment from his having those reasons to his doing A (he might have acted on the other considerations which include his feeling of moral obligation instead) and more importantly in that this complete explanation of the man's action in terms of his reasons does not have to be supplemented by a law-like generalisation to the effect that anyone (or perhaps just he) always acts thus when he has this particular combination of reasons for performing and for not performing an action of type A. (103) Conceivably, such a generalisation might be available, but one does not have to produce it to be said to have provided a complete explanation of the man's action in terms of his reasons. The explanation one offers, his having certain reasons for performing A, is one that is compatible with his having not performed A, and yet the gap that this seems to leave is one that one does not have to close.

How can an action which has been completely explained in terms of the agent's reasons for performing it be intelligible, in view of the fact that such an explanation is compatible with his having done something else? There are three extreme positions on the issue which ought to be avoided:-

(a) That a complete explanation of the action in terms of the agent's reasons ipso facto renders the action intelligible. This will not do firstly, because one might find the agent's reasons unintelligible; one might see no sense in his desires, fears or moral values. And secondly, and more importantly for our purposes, one might find his reasons for performing the action perfectly intelligible but be astounded by his acting on them in the particular circumstances. For one might know that he had reasons for not performing the action and find it incomprehensible that he should have allowed them to be outweighed by the considerations in favour of the action. A biographer of some historical figure might deem it incredible that his subject should have sacrificed the lives of a group of people in order to achieve a political objective. It may be that he does not assume that his subject did not want to spare the lives of these people, so he is not faced with the problem of how someone could be so callous as to be indifferent to the fate of his victims. Rather he is faced with the problem as to how he could be so callous as to subordinate the question of their fate to his further plans. It is not the man's values alone to which he takes exception, so much as their relative ordering. He might agree with his subject on the considerations to be taken into account as regards the political issue, but disagree fundamentally as to priorities.

(b) That one only finds an action intelligible in so far as one agrees with it. This is basically a reaction against (a). To be plausible it would have to be modified so as to take account of the agent's beliefs. For one might find an action intelligible, but disagree with it, because one knows of facts of which one believes the agent ignorant. So one would have to say something like, "one only finds an action intelligible in so far as one agrees with it, or would have done so had one had the same beliefs as the agent." But even this seems too strong: surely one sometimes finds the action of another intelligible, even though one in no sense agrees with it, nor does one think one would have agreed with it even if one's factual beliefs had been in accord with his. What is certainly true is that one sometimes finds the action of another intelligible, when one believes that one would not oneself have made the same choice in the same circumstances. This notion of what one believes one would have

done in the same circumstances is a little clearer than that of agreeing with a course of action. But the two notions are not co-extensive, and it may be possible to define a sense of "agreeing with an action" in which, if one knows of agent's reasons for doing something but does not agree with his action, one still finds a residue of unintelligibility in his behaviour.

(c) That one only finds an action intelligible in so far as one can explain why it was performed rather than some other. Unless one can find an explanation of why certain desires, fears, moral values etc. triumphed over others with which they were in conflict, one cannot claim to understand the action. This position undercuts the dispute between (a) and (b): to understand an action it is sufficient neither to know all the reasons for which the agent performed it, nor to be in any sense in agreement with him about it. The latter requirement is presumably not a necessary condition either. The trouble here is that it is not clear that (c) is really a requirement for the intelligibility of the action. The action of another can often seem intelligible even when we do not know why he acted on the considerations he did rather than others. And furthermore, if the biographer in (a) is presented with an account of why his subject acted as he did which does explain why political considerations outweighed more immediate humanitarian ones, it is not obvious that the action is made more intelligible. A gap in the explanation is closed, certainly; but are we to say that the action has been rendered intelligible to the biographer? If we do, we, at least, seem to be using a different sense of "intelligibility". But why not say that the biographer now has an explanation of an action that is unintelligible to him: that there are explanations of unintelligible actions as well as intelligible ones?

There is, I suppose, a fourth position: that an action has been rendered intelligible when the reasons for which it was performed have been given and it has been said that his action accorded with his preferences, that the considerations which triumphed over countervailing ones were expressive of his preferences, the relative strength of his desires, fears and moral convictions or of what he judged it best to do. But as we have seen this either degenerates into triviality, or amounts to the requirement that the action must accord with his (previous) judgements as to his preferences etc. But that this latter requirement is satisfied is what is normally to be expected; the man's behaviour would lose in intelligibility if it were not. The view amounts to position (a)

and is exposed to the same objections; it differs from the (a) in making explicit a possible requirement for the intelligibility of action, but this addition does not render it acceptable.

The three positions seem to approximate to three ideals of explanation: the first respects the fact that an explanation of an action is basically the presentation of the agent's reasons for acting as he did, and when this has been done there is a sense in which the explanation is complete. Any attempt to go beyond this would be to do more than to attempt to explain his action by reference to his reasons. It would be to deny the adequacy of the sort of explanation that is normally given of actions, in particular the sort of explanation an agent normally gives of his own actions. The second takes note of the fact that an action can be fully explained in terms of the agent's reasons, and yet seem puzzling. One might find it unintelligible or think it unjustified. And it is hard to avoid the impression that when one does have this sort of reaction to someone else's action one is craving for some further explanation of his action. But even if it is not further explanation for which one is looking, it still seems that something is missing, in the action or our understanding of it. And in this situation one naturally applies to the action such adjectives as "incredible", "incomprehensible" and "unintelligible". (104) Here we have the ideal of empathy asserting itself as a requirement of the understanding of action. The third position is close to the ideal of scientific explanation. If one cannot explain why an action was performed rather than some other, how can one claim to have explained it? To explain why an action was performed, admittedly involves giving the agent's reasons for performing it. But to do only this is not to explain why the action was performed rather than some other, if the agent had reasons for performing some other action instead. (105) Therefore a complete explanation of someone's action in terms of his reason for performing it falls short of a complete explanation of his action.

Now, as we have seen, all these positions are too extreme to take full account of the scope and limits of our understanding of people's actions. We are not always satisfied merely by the knowledge of the reasons a person had for acting as he did, nor do we always require that such an account must be supplemented by an explanation of how it was that the man's reasons for performing the action outweighed

the reasons he had for not performing it. The situation with regard to position (b) is rather less clear: one does not necessarily fail to understand someone's action just because one does not agree with it or because one would not have chosen as he did in the circumstances. It seems that I can understand someone's performing some action for monetary gain, rather than refraining because of his moral convictions even though I do not agree with his action (perhaps I condemn it morally) and do not believe that I would have acted similarly. (I am assuming that we both have the same relevant attitudes, a desire for monetary gain and certain moral convictions. It is not their existence but their efficacy in leading to action that is subject to my self-righteous scrutiny). But the notions of agreement and what one would oneself have done cannot be dismissed as simply irrelevant to the understanding of someone else's action. It is quite likely that I understand the man's acting for monetary gain in the model of my own moral failures, in dissimilar circumstances perhaps, but involving my acting on a desire for personal gain rather than in accordance with my moral principles, refraining from the action. It is also possible that I think that, had the moral issue been a little less serious, I would have agreed with the man's acting for monetary gain.

The converse of this is that one's failure to understand someone else's action even when one knows his reasons, often stems from one's being in complete disagreement with him as to priorities, our biographer was in such a position. So the notions of agreement and what one would oneself have done do seem closely bound up with one's understanding of the actions of others, although there is no simple entailment of the form: if one does not agree with someone's action (or does not believe one would have done the same in the circumstances), then one cannot really understand it. The point that one understands the behaviour of another by reference to one's own behaviour and standards can be developed further. For it were true that one could only understand someone's action if one agreed with it or would have done the same thing in similar circumstances, then it is difficult to see how one could avoid applying this principle to one's own past behaviour. No doubt there are cases where one is completely out of sympathy with one's own past behaviour, and although one can remember why one performed certain actions, one finds the actions unintelligible in the same way as one might find those of someone else unintelligible even when one knows the considerations that weighed with him. But this is not usual; normally one can understand one's own past actions even if one now condemns

them or would not now do the same thing. In this one is helped by one's memories of what it was like to feel as one did, to have the priorities one had. This ability and the resultant awareness of the changes that one's own scales of values have exhibited probably helps one to find intelligibility in the actions of others when one does not agree with them, or would not have done the same.

The application of (b) to one's case, then, shows fairly clearly that any simple formulation of it is likely to be too strong. A similar application of (c) to one's own case would have extremely unpalatable, if not actually incoherent, consequences. Our understanding of our own actions, even those we are at present performing, would, on this view, be severely limited. Clearly the type of explanation usually appropriate to one's own actions is that of simply giving the reasons why one acted as one did, or is acting as one is, <sup>that is, one's reasons</sup>. And similarly, one's understanding of one's own actions increases in proportion as one learns the reasons why one acted (one's understanding might be defective in that one might be ignorant of unconsciously operative motivational factors). With the actions of another, and to some extent of one's past self, the position is more complicated; as one's comprehension might be baulked by the kind of failure of <sup>c</sup>sympathy discussed earlier. This, together with the traditional scientific demand for explanations which tell us why someone performed one action rather than another when he had reasons to perform another action instead, should give pause to anyone who is inclined to think that explaining actions in terms of the agent's reasons for performing them exhausts the role of explanations in the realm of action. (106)

Much classification in this area could, of course, be achieved by labelling the various kinds of explanation, of understanding and of intelligibility. Here I have only attempted to indicate the diversity of these concepts by describing three unacceptably extreme positions each of which emphasises one kind of understanding to the exclusion of others. It is at least as important to see the connections between the various types of explanation and understanding. To understand a person's action, it is not always sufficient to know the reasons why he acted as he did, but at least it is necessary. The empathetic understanding of a person's action presupposes that one knows the reasons for acting as he did, and also any reasons he had for not acting in that way. If one is wrong about his reasons, one's understanding is vitiated. And if one wants to know why someone chose to do A

because he wanted X rather than B because he wanted Y, one is posing a question framed in terms of the agent's reasons for acting. To what extent one can always expect answers to questions <sup>of this kind</sup> framed in this way is not obvious: neurophysiology can explain why a certain bodily movement occurred in a way that shows why no other bodily movement occurred but can this be translated into an explanation of why the man acted on one desire rather than the other? Much will depend on the possibility or actuality of correlation between mental and neural states, and maybe on the possibility of finding some measure of the "strength" of a desire from a knowledge of which it can be predicted which of several desires ~~is~~ in conflict will lead to action.

Not only do some ways of understanding action presuppose or supplement others, they might on occasion substitute for each other. Thus, if our biographer is given an explanation of why the figure he is studying acted as he did, which explains the power of his political convictions in terms, say, of childhood experiences, <sup>of</sup> in the comparative weakness of humanitarian sentiments of a more personal kind in terms of, say, brain-damage, he is not likely to feel cheated. On the other hand, this will not help him much to empathise with the man. And he might still go on feeling this lack, as it is not clear that the explanation he has been given places the politician beyond the range of empathy, as an explanation in terms of more radical psychological disturbance might. Still his blank incomprehension of the man has been to some extent mitigated. He is making do with one kind of understanding, as he lacks the other. Perhaps any conscientious biographer feels the need to understand, in the sense of "to empathise" with his subject (here is a good illustration of why it is too simple to equate this kind of understanding with agreeing with his actions, or thinking one would have done the same: few of us would require <sup>of</sup> a biographer of Hitler that he should agree with Hitler's actions) but such understanding may be denied him.

So far we have been concerned with the contemplative aspect of our responses to another's actions, with our explaining or understanding those actions. But people also hold other people responsible for what they have done, feel gratitude towards them or harbour  $\neq$  resentment against them, and such attitudes lead them to take action. (107) Now when one views an action in this light one is always concerned with the fact that the agent did one thing rather than something else, where this "something else" might merely

be refraining from the action in question. Let me develop this point with reference to a specific example.

Suppose someone hits me and I resent this. Now certain conditions must be satisfied if my resentment is to be in order. Firstly, his hitting me must be intentional, or if it was not intentional, at least I have to think of him as being culpably negligent (if, for example, he was trying to hit something or someone else and hit me by mistake). If the injury he inflicted upon me had been by a more circuitous route than simply hitting me, there would have been another possible situation that would have justified my resentment, namely where the injury inflicted was neither clearly intended nor unintended by him, but simply accepted as a necessary consequence of some intentional action; when, that is, his action of injuring me was voluntary rather than intentional. Now there are other circumstances that might tend to inhibit my resentment, or at least ought to. One of these circumstances is his having good reason for hitting me: for example, he is acting under duress or possibly I have unduly provoked him. If my resentment is inhibited in this way, it might be that I assume that he wanted to refrain from hitting me but that this desire was overridden by other considerations, and I regard this as justifiable in the circumstances (whether I must assume that he had some such countervailing desire, that other considerations at least weighted with him even if they were overridden, I am not clear).

On the other hand, if my resentment is in order, I do not have to assume that he had any countervailing desires or that he seriously entertained any considerations in favour of not hitting me. If I learn that someone hit me either out of sheer malice, or because of some further design and that he had nothing against hitting me, no desires or values which weighed against hitting me, my resentment is unlikely to be diminished; in fact, it will probably be increased. This suggests that the cases where I can justifiably feel resentment towards someone might not be coextensive with those where I ~~can~~ clearly state grounds for holding him responsible. For it is far from obvious that I can hold him responsible for his not having certain desires or values, still less that I must do so if I can justifiably feel resentment. And given that he had no desires or values which conflicted with the desire to hit me, how can I require of him that he should not have acted on his desire to hit me? Admittedly, in the first chapter, it was argued that there is no entailment from an unopposed desire to action; but this



can hardly be used as a basis for my demanding that the man should not have acted on his desire to hit me. For it is expected that an unopposed desire normally leads to action; cases where it does not are bound to seem irrational or even pathological. To say that he ought not to have acted on his unopposed desire to hit me is tantamount to saying that he ought to have behaved irrationally, or that the normal processes of agency ought to have broken down.

Such a conclusion is unpalatable. One way in which one might attempt to avoid it is by developing the claim of Nagel's that a belief about another's needs or desires can itself motivate without the need for desires on one's own part to explain the motivation. Thus given that the man believes that his hitting me will injure me, or that I do not want to be hit, he has a reason for not hitting me. The situation is then one of conflicting considerations, and I do not have to say that he ought not to have acted on an unopposed desire. If the man had not believed that his hitting me would injure me, then the possibility would be opened that any resentment against him was not justified. This suggestion seems to do the work required of it. On the other hand, it is hard to assess the extent to which it involves a certain a priorism. After all, the man who hit me may have experienced no conflict, and might not allow that the fact that his action would harm me was any reason at all for his not performing it. Perhaps, if he consistently took this attitude to the interests of others, he would reveal himself a "moral idiot" or "psychopath", and thus outside the range of normal human responses and reactions. (108) Such a man it should be noted can be described both on Nagel's theory and on that of desire-based motivation. On Nagel's view the "psychopath" does not allow that another's interests provide any reason for acting in one way rather than another and in this he (the "psychopath") is presumably wrong. On the theory of desire-based motivation, the "psychopath" is one who never has any desires to further another's interests or to refrain from harming them. (It is not sufficient that he should have no general desire that others, whoever they may be, should not be harmed as this may be true of many people who are not psychopaths: they often have desires for the welfare of specific persons and act on them).

Nagel's view, then, smacks of a priorism, although it must be admitted that the considerations I have adduced to show this, smack, themselves, of introspectionism, an equally suspect methodological procedure. Rather than opt for Nagel's view, therefore, I would

prefer to leave it an open question as to whether there might be cases where one can justifiably feel resentment, even though one would be hard-pressed to say just in what one's ascription of responsibility consisted. In the above example, we could hardly on the basis of this incident alone, conclude that the man who hit me was a moral idiot, an inappropriate object of reactive attitudes. In fact, as mentioned earlier, resentment is likely to be particularly strong.

If such cases are possible, I do not think that this tells in an obvious way for or against any particular theory of responsibility. For if one assumes that to be justified in feeling resentment, one must be able correctly to ascribe responsibility to the agent for what he did, one is going to have difficulty in saying how this is possible, whatever theory of responsibility one adopts. Both the libertarian and compatibilist-determinist will find it hard to avoid giving the impression that they are either holding a man responsible for not having certain desires or values (109), or that they are blaming him for acting on the only consideration which mattered to him in the circumstances. Whether notions of "contra-causal freedom" or the "notions of his having been able to do otherwise, if he had chosen," are employed, it is difficult to see how they can be applied to the type of case in question. They will both have to say either that resentment is appropriate sometimes even when the ascription of responsibility has no obvious basis or describe these cases where resentment would very naturally be felt, as not justifying resentment. And the same would apply if we alter the cases slightly, and enquire/<sup>not</sup>as to the status of resentment, but as to the status of moral indignation on the part of others and the practice of punishment by the authorities. When the man who has no desire to do otherwise hits me, not only would I naturally feel resentment, but others might feel moral indignation on my behalf and possibly the law might step in to punish him.

I said above that when we feel resentment we are concerned with the fact that someone did something rather than something else (unlike explaining someone's action by giving his reasons for performing that action, where one stops short of explaining why he performed that action rather than some other which he also had reason to do). Because of the possibility of such problematic cases as the one just discussed, it is not clear that this amounts to saying that in feeling resentment, we are concerned with the fact that someone was moved by certain of his desires, values etc.

rather than others. But normally, when we feel resentment against someone we can express our complaint in a way which does bring to the fore an assumption that he acted on certain considerations rather than others, and where these considerations are desires, values etc. which we attribute to him. If Nagel is right, perhaps we always can. Thus if someone failed to keep an appointment with one and his excuse is that he was tired, I might say, "That is not a sufficiently good reason. You ought to have acted on your desire to keep your promise, or on your moral conviction that it is wrong to break promises". Here it may be objected that the man may have had no desire whatsoever to keep his promise and no moral convictions that it would be wrong to break a promise. This is true, of course, and now our example begins to move in the direction of the one discussed above, where the man desires only to hit me and has no reasons for not doing so, unless we can save it by means of Nagel's thesis.

Now, if I am right that when we feel resentment we are normally concerned with the fact that someone acted on certain considerations rather than others, all of which we assume to have been in some sense his, then it looks as though resentment enters at the point where the explanation of his action in terms of his reasons ends. Once we know the reasons which the man had for his action the stage is set for us to feel justified or unjustified in harbouring resentment. If this is all there is to the matter then our digression on the subject of whether the person against whom we feel resentment, had any reasons for not performing the action, was somewhat irrelevant; we could have arrived at this conclusion without it. But, in fact, we do take into account the question as to whether someone had any reasons for not performing the action we resent. This is shown by our uneasiness about the case of the man in a certain situation who had no desire other than to hit me, and our readiness to think of him as removed from the sphere of ordinary interpersonal attitudes if his action is typical. It is shown more clearly by the fact that sometimes we let people off the hook if we are convinced that they sincerely believed they had no reason to do otherwise than they did, if for example they subscribe to moral principles different from ours. Perhaps the cases of the man hitting me and the man failing to keep his appointment do not furnish particularly plausible examples here. But suppose I recount to someone an experience of an obviously personal nature, which he promptly relates to others. His excuse is that I did not place him under any obligation not to tell before beginning the disclosure; I did not state that

my experience was being told in confidence. If I come to think that he sincerely believes that only when the disclosure is clearly solemnised as in confidence, is he under any obligation not to tell it, then my resentment might well be mitigated - not eliminated, perhaps, as I might still feel that he acted in crass disregard of my feelings. I might also think his view a stupidly legalistic one, but that is a different line of criticism.

With gratitude, the reactive attitude which is in some ways the converse of resentment, the importance of countervailing reasons is clearer still. If someone confers on me a benefit, and I believe he did it intentionally and with the object of helping me, then, on this account, alone, I ought to feel gratitude. But if I learn that he had good reason for not acting as he did, that his action involved the frustration of some of his own desires, then (assuming his self-denial is not <sup>so</sup> extreme as to begin to look ~~irrational~~ ~~unnatural~~) I have reason to feel more gratitude than ever.

So one feels justified resentment or gratitude to someone when one knows the reasons they had for acting as they did, and possibly any reasons they had for acting otherwise. But one may not know why certain considerations moved them to act rather than others. Indeed, some explanations might tend to inhibit resentment. But why is one justified in feeling resentment when one has no explanation of why he acted in one way rather than another (when one's explanation is only of the type we have called a complete explanation in terms of his reasons for performing the action)? Surely, behaviour which one cannot really explain can justify no other attitude than bewilderment. (110) And yet such situations are just the ones in which we do feel resentment - when we have a complete explanation of an action in terms of the agent's reasons for performing that action. Let us approach this problem by asking simply, "Why is resentment justified?" without any implication as to the source of any scepticism we might entertain, and by asking the related but distinct question, "What is the point of resentment?"

The fact that a person has behaved in a certain way can make it rational for me to take action towards him. In the light of purely causal considerations I might attempt to influence his behaviour so as to prevent a recurrence of the incident. Or I might simply assume that what he has done once he is likely to do again, and so avoid him in future. But such policies could be adopted without my feeling resentment, or, indeed, if I did not

regard him as a proper object of resentment: if, for example, I considered him to be mentally deranged.

Strawson (111) has argued against the view that it would ever be rational to abandon our humane concern with other human beings expressed in our adopting reactive attitudes such as resentment and gratitude even if it were known that the thesis of total physical determinism were true. Such a rejection of our reactive attitudes would, According to Strawson, if I understand him correctly, neither be rational nor possible. But it is not easy to see that gives sense to our reactive practices in the first place. Consider a class of acts typically performed out of resentment, those of revenge, (not all acts performed out of resentment are acts of revenge: admonitions, for example, are not, yet they may be motivated partly or wholly by feelings of resentment). Kenny (112) has argued that an act of revenge is performed with the intention that one's act should exemplify a certain pattern; it is not merely an act that exemplifies a certain pattern. He also warns against taking either of the extreme views that revenge in no way serves the interests of the avenger, or that it always does. Now what is relevant here is not so much whether the action serves the interests of the avenger, as whether he performs it in order to serve those interests. And, as far as I can see, if someone performs an action which injures someone else who has injured him, and he does this solely to further his own interests, he is not performing an act of revenge. Such an act could be performed by someone who felt no resentment, which is not to say that the words "because he injured me" would not occur in his description of his reasons for acting. The simple elimination of someone who has injured me, in order to prevent an occurrence, might be such a case.

It does seem that there are acts of pure revenge, where a person because of some harm he believes another to have voluntarily done him, acts<sup>so</sup> as to harm this other, even though he does not expect any good to accrue to him from this. No doubt most cases are mixed; considerations of one's own interests play a part. But insofar as someone allows considerations of his own future interests to influence his behaviour to someone, his behaviour is to that extent less purely revenge-behaviour. If he says, "I'll teach him a lesson!" and means it, then it looks as though he wants more than just revenge. He wants someone to realise the error of his ways, or to refrain from certain actions in the future. However, if I am wrong in this, and we do not think of considerations of one's

own welfare as detracting from the purity of one's revenge-seeking, the cases I have called acts of pure revenge, at least exist, ~~and~~ are paradigmatic. And it is difficult to give an account of them which makes this seem at all rational.

A natural thought at this point is that any attempt to do so is misguided. The whole point of revenge is that it is action for a certain kind of reason, and any attempt to exhibit or deny its rationality will involve an illicit attempt to assimilate it to action for some other kind of reason. It will be of a piece with such enterprises as explaining or justifying altruism in terms of self-interest, morality in terms of prudence or induction in terms of deduction. (113) Perhaps it will be said that the practice of revenge is a "form of life". At any rate, revenge is sui generis and cannot be rendered intelligible in terms of actions the reason for which is of some other kind.

That revenge is sui generis is, of course, the burden of my argument concerning Kenny's views. But can one on this basis regard it as rational? Or rather can one make the more subtle move of saying that although a general practice such as revenge can be neither rational nor irrational, justified ~~nor~~ unjustified, particular actions can be shown to be rational or justified by reference to it? There is no telling what absurd behaviour might get a "justification" in this way. In any case, many people will feel that although resentment is justifiable, revenge is not, because the latter, although showing the autonomy of the former, is subject to veto by certain moral principles. Revenge, on this view, would be intelligible but not justifiable. So let us return to resentment.

If one asks for the point of someone's resentment on a particular occasion, it is obscure what is being asked. It might just be a request for the reason for his resentment. But if it means: what is good about his having a feeling of resentment, there may be no answer to this question. And if one thinks it is bad to feel resentment (either on certain sorts of occasion or generally) as distinct from <sup>to</sup> ~~act~~ act on it, it is not obvious what one can do about it. We do not suppose that feelings of resentment are directly subject to the will. The justification of resentment is sui generis; the reasons one gives for one's resentment are of a different kind from those one gives for one's gratitude. More importantly, they are of a different kind from those one gives for one's action: one does not say what is good about one's harbouring

resentment, whereas one does say what is good about one's actions. Strawson seems on strong ground when he maintains that it would not be possible for human beings to abandon reactive attitudes, even if some general theoretical conviction should seem to undermine their basis. (114)

Still this does not allay one's doubts as to the rationality of resentment. Resentment might be inescapable but why should it not also seem a burden, a perpetual affront to the rationality of Man? Even the consideration that it is wrong to ask for the point of resentment seems a little slick. For one can act on one's resentment and actions are something for which one can demand the point. If the action is one of revenge, and the rationality of this is questioned, one has nothing more to offer than the "sui generis" retort. And if, as many of us do, we regard acts of revenge as wrong, we seem already to be treating resentment as suspect, suspect that is as a motivational influence. If a moralist were to go further and maintain that any show of resentment in action, say an admonition, was wrong, he would seem to be engaged in all out attack on resentment, tempered only by a reluctance to regard resentment itself as subject to the will. His view would be that resentment should never be manifested in action; it ought never to be more than a mere experience.

If then I have a feeling of resentment towards someone I believe to have intentionally injured me, over and above my adopting a policy designed to prevent a recurrence of the incident, there is a question as to the rationality of this feeling. Ought I, for example, to regard it as completely unreasonable, in the same light as I might regard<sup>it</sup> if I believed it to have been induced by a drug? A sceptical attack on the rationality of such a feeling could be built on several considerations, other than those already mentioned:- (a) Why should what a person has done provide any reason for an attitude to him now? Is there anything about the person at present which justifies my attitude? Does something of his past deed endure in him, or does his past deed alone justify my attitude, without the need for anything to linger on and mediate between his deed and my attitude? Certainly, if I think he is likely to repeat the performance, I can take steps to avoid being the victim in future, but such preventive behaviour could be indulged in without my feeling resentment and if I did not think him a proper object of resentment. We are in danger of allowing the justification of resentment for a past deed to collapse into the justification of any

manipulatory behaviour we might indulge in to prevent a recurrence of the incident. It is not clear that the fact that a person's actions in the past may provide a justification for certain behaviour designed to prevent his behaving similarly in the future, can be used to justify resentment. The considerations seem too disparate, and the "sui generis" defence of resentment and revenge emphasises this.

(b) Why should the attitude of resentment be directed at the person one believes to have intentionally injured one? It is easy to see why one's behaviour designed to prevent a recurrence of the incident should be directed at him. But if true resentment and revenge have no further point, what gives sense to their directedness? Suppose we encountered a tribe who dealt with crime and the criminal in the following way: they engaged in behaviour calculated to prevent him from performing similar actions or in behaviour calculated to deter him and others from performing such actions in the future. They also had feelings of ~~honour~~<sup>honor</sup>, resentment and moral indignation at the crime. These latter, however, were not directed at the perpetrator of the crime, but elsewhere, say at his father or at the gods or at someone earmarked by the tribal wise-man. Their manipulatory behaviour that was designed to secure future benefits or prevent future harm was completely divorced from their "intuitions of fittingness". Admittedly, we might be reluctant to translate certain words in their language as "resentment", "moral indignation" and so forth without further comment, but it would be fairly obvious that certain practices and attitudes in some ways played the same role in their world as did resentment, revenge and indignation in ours. (115) What is more, members of the tribe could justify their behaviour by the "sui generis" move. The conclusion to be drawn from this fantasy is again basically a question: how can the fact that behaviour designed to prevent a repetition of someone's action (by him or others) should rationally be directed at him in most cases, be used to justify the directedness of resentment or indignation?

If one attempts to overcome these difficulties by insisting on the causal efficacy of behaviour designed to influence the conduct of those who have committed or might commit a wrong, in the manner of utilitarians and some of those who wish to show that determinism and freedom are compatible, one runs into certain well-known difficulties in giving a reasonable account of justice. (116) But furthermore one ceases to justify the behaviour as punishment or revenge. One can show that a certain way of treating those who have



wronged one is rational, but one does not succeed in justifying it as revenge. And one can show that a certain way of treating wrongdoers is rational, but does one thereby succeed in justifying it as punishment? Or does one justify an act which might have been performed as a punishment but by giving a reason for it which is not that appropriate to punishment? The utilitarian approach, and I do not claim to be able to offer anything better, is really a denial of the rationality of any of our behaviour and attitudes that go beyond manipulatory practices.

Certain problems arise about reactive attitudes quite independently of any conflict there might seem to be between their adoption and a belief in physical determinism. I have tried to show this, at least as regards resentment and behaviour motivated by it, by arguing that there is already some conflict between resentment and commonly held moral beliefs and by arguing that the directedness of our feelings of resentment, at someone for his deed in the past, seems to hang in the air unless a justification is imported from the seemingly disparate sphere of the utilitarian manipulation of that person. (117) It will doubtless be felt, and with some justice, that I have been implicitly assuming a non-human stance towards resentment, perhaps that of a pure egoist, or a pure altruist or someone, possibly a utilitarian, whose attitude spans these two positions, who feels that an action can only be justified in terms of the good expected from it either for oneself or others. Such a "person", if any there be, would obviously fail to see any sense in such behaviour as revenge.

I think that drawing attention to the moral convictions which many of us already possess and which are in direct conflict with our feelings of resentment, together with a consideration of the inadequacy of the "sui generis" argument can go some way towards meeting this objection. But there is a good point implicit in the objection which can be brought out in the following way: not all the measures we take against those who injured us are actuated by feelings of resentment. If someone hits a man in a public bar, the latter might simply stay away from the place in future. Or he might get the landlord to exclude the aggressor. The first course of action need not be motivated by resentment: he does not have to be trying to get his own back at the aggressor, by denying him his company. Nor is it true that the second course of action must be actuated by resentment. Still less does the judge who applies the law necessarily have feelings of outrage or moral indignation at

those whom he sentences (here we are considering what Strawson calls the "vicarious analogue" of resentment (118) ). How different from would a person be whose behaviour was never actuated by feelings of resentment or moral indignation? His behaviour need not be very different, for there is considerable overlap among people like us, between revenge and punishment on the one hand and measures of prevention and deterrence on the other. His inner life, one would want to say, would be very different.

The man who was assaulted in the public bar, with cool deliberation, arranges for the aggressor to be excluded henceforth so as to prevent any repetition of the incident. He disclaims any feelings of resentment, maintaining that having him banned from the bar was no more an act of revenge than if he had decided to stay away himself. One might have some difficulty in accepting his account; but the difficulty is not insurmountable. What would be very odd, however, would be the claim that all his behaviour was of this calculating sort. And yet the behaviour itself might be very similar to that of the others. It is the description of his inner life that seems incredible. Wittgenstein's example of the tribe who regard their slaves as machines springs to mind here; much of their behaviour might be similar to those of a tribe who regard their slaves as human beings. But perhaps not all: there would be no averting of the eyes when a slave was seriously injured. (119)

But with our example the behaviour of those who feel resentment, and those hypothetical beings who do not, might be even more similar. For those who do might think it wrong to give way to feelings of resentment, or in those cases where revenge serves no utilitarian purpose in addition, it might usually happen that their desire for revenge is overridden by other considerations. There is, of course, a big difference between someone who does not have certain desires, emotions or feelings and someone who always keeps them in check. And yet there may be no clear indications as to which category an individual belongs in his behaviour; or if there is, it might take the form of behavioural evidence for his being subject to various countervailing considerations.

Wittgenstein asks/<sup>us</sup> to try to imagine that someone who is injured and is writhing on the floor, is not in pain. (120) Let us alter the example: we admit he is in pain, but now let us suppose that he believes that it is we who have injured him. And suppose he later takes steps to injure or incapacitate us. Now try to imagine

that he feels no resentment and that his behaviour is designed purely to prevent a recurrence of the incident. The difficulty of doing this seems almost of the same order as that of performing Wittgenstein's thought-experiment.

On the other hand we can imagine facets of personality and situations which might decrease someone's liability to feel resentment (not just his liability to give way to resentment). A certain kind of feeling that one is superior to other mortals or a Spinozistic resolve to see others as part of the natural order might have this effect. As might a conviction that one's misfortunes are always to some extent one's own fault in that one has failed to assess another's character adequately or that one is too vulnerable to injury or that one is impotent to defend oneself. Perhaps even a high frequency of injury by others might affect someone in this way. As I said earlier, it is hard to imagine a magistrate who regularly doles out penalties experiencing feelings of moral indignation as regards each and every case with which he deals. (121) Similarly someone whose social (or non-social) environment came to approximate a Hobbesian "war of all against all" might possibly cease to resent the actions of others, whatever measures he might take against them. That such a change of attitude might be better explained in terms of the weakening of the notions of reciprocal obligation in such a situation does not alter the fact that the possibility, if it is a possibility, constitutes one of a diminished liability to feel resentment.

To sum up: such reactive attitudes as resentment and behaviour motivated thereby seem part and parcel of normal human nature, or better of the human condition, as certain very general <sup>social</sup> ~~social~~ facts might also be involved. If we wish to regard them as rational we have to regard the type of rationality they manifest as *sui generis*. Alternatively, we can attempt to evaluate them morally; we might regard resentment ~~and~~ gratitude as fundamentally ~~arational~~ or even irrational, but attach a different moral significance to them. Resentment one might condemn; but gratitude one might value. It is harder to imagine a moralist condemning gratitude than it is to imagine one who indulges in a wholesale condemnation of resentment. But perhaps it is not impossible: someone of an extremely altruistic bent might think that feelings and acts of benevolence are owed to all human beings, and that those who have benefitted one are in no special position in this respect. Perhaps he might add that someone who has benefitted one was, in any case, only doing his duty. (122)

Strawson seems right to hold that we could not "turn off" our reactive attitudes, merely in the light of a general theoretical conviction of the truth of determinism. But there is more to the matter than this. The rationality and morality of resentment and revenge can be attacked from a number of positions which have nothing to do with determinism (and certain attitudes of mind would seem to diminish one's liability to such feelings as resentment). Although it is not to be supposed that all these criticisms can easily be transferred to the case of other reactive attitudes, such as gratitude, some of them can, and so there is a general problem as to the status of such attitudes, which is prior to that of how their status would be affected by our coming to believe in the truth of the deterministic thesis.

The account of action that has been presented has been a causal one. The statement that someone acted as he did because of certain fears, desires or moral values requires for its interpretation that we give a causal force to the "because". Now if the man had other desires, fears or moral values which conflicted with those on which he did act, we can ask why he acted on the considerations he did, thus raising a question which goes beyond any that can be answered by giving a complete explanation in terms of his reasons. It is far from obvious that such a question is illegitimate, and a consideration of the puzzlement we can feel about someone's acting on certain considerations rather than other conflicting ones, and the sorts of account which would mitigate our puzzlement seem to suggest that such questions are not merely prompted by an unquestioning adherence to a deterministic presupposition as to what the world will turn out to be like. But if such questions ~~could~~ ever have an answer, it is hard to see how the answers could fail to involve causal considerations, considerations as to how it came about that a certain desire caused a certain bodily movement instead of it happening that another desire caused another bodily movement.

When we come to reactive attitudes we find we are typically concerned with the fact that someone acted on certain considerations rather than others, not the explanation of his acting on certain considerations rather than others. But certain types of explanation

can make resentment, for example, seem unjustified (he was given a drug which is known to make one's sexual desires uncontrollable, let us say). The fear of determinism, in one form at least, might be expressed as the fear that there might be explanations of all actions, that are explanations of this sort. And a causal analysis of action could very probably strengthen this fear (this is not to suggest that a non-causal analysis, whatever such an analysis which tried to account for the force of the "because" as demanded earlier, might be like, would circumvent the problem of determinism. For if the bodily movement was determined, this might be sufficient to pose the problem (123).) But if a causal analysis is correct, as has been argued, then this may mean that there lies latent in the concept of action a factor which could undermine the rationale of our attitudes to the actions of others, though not, of course, much of our behaviour towards them, such as that which I have labelled "manipulatory".

My aim in this study has been to give an account of acting for a reason and a general discussion of freewill and determinism is beyond its scope. Nevertheless, the discussion touches on the problem at various points: the causal theory of action, the need for explanation that remains after a complete explanation in terms of an agent's reasons has been given and the significance of an agent's being moved to act by certain considerations (rather than others) for our attitudes to him. As the interest in the philosophy of action stems in large part from an interest in this problem it is worth indicating these points of contact. I would like to end with a plea that the issues at the points of contact be discussed also on their own merits: that the causal theory of action be not judged on the basis of its association with what is to some an obnoxious metaphysical doctrine; that the question why someone acted on the considerations he did rather than on considerations conflicting with them, be recognised as a possible question; and that the problem of justifying reactive attitudes and behaviour resulting from them which arise prior to any clash with determinism be given due consideration.



Footnotes

1. G.E.M. Anscombe "Intention" § 36
2. At this point my treatment of this issue begins to merge with that given by Goldman (p.111 et.seq. of "A Theory of Human Action"). He writes "To say that S imbibes poison does not entail that S dies soon thereafter. However, together with the further provisos that S does not take an antidote, that S has not developed an immunity to this poison etc., it perhaps does entail that S dies shortly afterwards. Even if we grant this entailment, however, it does not follow that S's imbibing the poison does not cause his dying." Goldman, who is not so much arguing against the ACP thesis as claiming that it does not rule out a causal connection between desires and action, is presumably thinking of the move of construing whatever accounts for a poison's failure to kill in a particular case as an instance of one of the provisos in the entailment being unfulfilled.
3. Raziel Abelson "Review of Richard Taylor, 'Action and Purpose' " pp. 183 - 184. This is the passage discussed by Goldman (see footnote 2).
4. Indeed the best statement of the ACP thesis that I can find occurs in the writings of D.F. Pears who is critically disposed to it. See "Are Reasons for Actions Causes?" in A. Stroll (ed.) "Epistemology", "Desires as Causes of Action" in "The Human Agent, Royal Institute of Philosophy Lectures Vol.1", "Two Problems about Reasons for Action" in "Agent, Action and Reason" ed. Binkley, Bronaugh and Marras.
5. The best-known statement of this argument occurs in A.I. Melden "Free Action". As Davidson points out in a footnote to "Actions, Reasons and Causes" Journal of Philosophy 1963, variants of this argument are to be found in Kenny's "Action, Emotion and Will", Hampshire's "Thought and Action", Winch's "The Idea of a Social Science" and Peter's "The Concept of Motivation" and that "in one of its forms, the argument was of course inspired by Ryle's treatment of motives in 'The Concept of Mind' ". I have not adopted for discussion any of these particular formulations without modification because, as will be evident from the text, I think the logical connection between desire and action has to be brought out in different ways to suit different cases and to rebut certain false moves which we might make in an attempt to refute the anti-causal theorist's contention.
6. I do not in this essay propose to discuss every possible

argument that might be brought against the causal theory of action. These arguments of which I am aware, and have not considered, I believe to have been adequately dealt with by Davidson, Goldman and Wilson. These include (a) the claim that reasons consist of such entities as desires and beliefs which ~~are~~ states and not events, and therefore not causes (b) the claim that we can cite a person's reasons for an action without being able to adduce any generalizations to the effect that certain desires etc. are always followed by certain actions, but that law-like generalizations are a necessary component of causal explanations; therefore explanations of actions in terms of the agent's reasons are not causal explanations. I discuss this point briefly in connection with certain claims of Thalberg, in the next chapter. (c) the claim that a person's knowledge of his reasons for acting is direct, non-inductive and possibly incorrigible and knowledge of causal relations cannot have any of these features. For treatments of (a), (b) and (c) see Davidson "Actions, Reasons and Causes" Journal of Philosophy 1963 (pp.693 - 695, 696 - 699, 699 - 700 respectively) and for a treatment of (a) see Goldman "A Theory of Human Action" (pp.86 - 98) See also Davidson "Causal Relations" Journal of Philosophy 1967 and Chapter 2 "Causal Relations" of "Emotion and Object" by G.R.S. Wilson for analyses of the notion of a cause adequate to meeting the above objections (there are certain differences between these two accounts but both seem successfully to show that objections (a), (b) and (c) are not well-taken).

7. I do not propose to discuss practical reasoning in any detail. But I would opt for a causal account for reasons which parallel those which suggest a causal theory of action (one wants to distinguish "He desires A because he desires B and believes A is a means to B" from "He desires A and he desires B and believes A is a means to B"). Such an account is defended by Goldman in "A Theory of Human Action" (pp.99 - 109). However, I do not think that if someone desires B, and believes that performing A is a means to getting B and for this reason performs A, then his action must be mediated via the generation of a desire to perform A. At least I do not know of any arguments that it must; it seems that the desire for B together with the belief might more directly lead to the action. My consideration of a view of Nagel's in the next chapter could be modified so as to cast doubt on the view that the two possibilities (that which involves mediation by a desire for the means and that which does not) are clearly distinct. I am unclear on Goldman's position on this issue.



8. I owe this way of putting the matter to Mr. B. O'Shaughnessy. The original version of the argument is to be found in Davidson's "Actions, Reasons and Causes" Journal of Philosophy 1963 p.688
9. Kenny "Action, Emotion and Will" pp.112 - 115
10. Kenny "Action, Emotion and Will" pp.115 - 116
11. Goldberg "Can A Desire be a Cause?" pp.70 - 72 Analysis (1965)
12. My thinking on this issue stems from J.R.S. Wilson's treatment in "Emotion and Object" of Kenny's claim that emotion and object are non-contingently connected. He writes of one interpretation of the view that emotions, but not sensations have objects that "it can't be an isolated fact about emotions that they are essentially directed to objects. If it were, we would have to be able to say of an emotion E and a sensation S : 'It happens that S has an object, but it differs from E in that it doesn't have to have an object'. This is impossible: modal properties can't exist in isolation". My treatment of the anti-causal theorists' claim that there must be descriptive matching if a desire is to explain an action, is to take seriously the fact that the "must" is followed by an "if" clause, and to ask how the requirement that there must be this matching if the explanation is to be of a certain sort is supposed to rule out the explanation being of a causal kind. As far as I can see there could be explanations in terms of the citing of a cause some of which in virtue of this requirement being satisfied, are explanations of actions in terms of reasons. If it is said that there is descriptive matching in explanations of actions in terms of reasons and although there must be if the explanation is to be of this sort, it is simply the fact that there is which rules out a causal connection, this makes it totally obscure why it is the type of matching appropriate to giving an agent's reason for his action which has this effect (that of precluding a causal connection).
13. It should be noted that the match is, for example, between "X is above temperature T at  $t_1$ ", and "X is above temperature T at  $t_2$ " and, of course, the completeness of the match is broken by the difference in the temporal specifications. The point is that one could define a kind of descriptive matching which was instantiated by two descriptions of states of affairs, and this fact does not affect their status as terms of a causal relation. Such descriptive matching is not of course that appropriate to that between an action and the reason for which it is performed. "X performs A" and "X desires to perform action A" match in a different way and thus differ in a different

way (one mentions a desire and the other does not). And once more we have to demand: why does the latter kind of matching rule out a causal connection? The statements "X is above temperature T at  $t_1$ " and "X is above temperature T at  $t_2$ " make mention of different states of affairs, it is true: but then so do "X performs action A" and "X desires to perform action A". It is only the state of affairs specified in the former statement and the object of the desire referred to in the latter that could possibly be thought to be the same state of affairs. And reasons have already been given why this is a shaky basis for an anti-causal theory.

14. I owe this point to Mr. B. O'Shaughnessy. If one performs an action involving a bodily movement, then the bodily movement will have causes which are events in the spinal cord. One's desire and belief will have to cause the bodily movement via such events.
15. "The Philosophy of Action" A.R. White (ed) p.7 of the editor's introduction.
16. Frankfurt, "Freedom of the Will and the Concept of a Person" Journal of Philosophy 1971.
17. Irwing Thalberg "Enigmas of Agency" Chapter 3 - originally co-authored by Arnold B. Levison.
18. op.cit. pp. 74 - 75
19. op.cit. p.75
20. op.cit. pp. 81 - 82
21. op.cit. p.85
22. op.cit. p. 86
23. Hence I believe that Thalberg's terminology of "essential explanations of action" is somewhat infelicitous, as to give those criteria we have for saying that an action is an action of a certain sort, need not be an explanation of the action in any obvious sense, and yet it is essential explanation of the action to Thalberg.
24. Note further the following remark of Kenny's in "Action, Emotion and Will" - "An executive who drops into a pillar-box a cheque to a blackmailer, and a love-letter to his mistress is performing at the same time two actions, one out of fear and one out of love; yet he need not be crying, starting, smiling, holding his breath, cooing, gurgling, or suffering visceral commotions, and indeed he can hardly be doing all these at the same time" (p.43) I do not know whether Kenny is correct to claim that we have here two actions, but that is not important to the point at issue. Indeed much of Kenny's treatment of fear in this book suggests that we can fruitfully compare desire and <sup>fear</sup>belief in so

- far as their capacity to motivate is conceived.
25. Davidson "Actions, Reasons and Causes" *Journal of Philosophy* 1963 pp.692 - 693
  26. Except to mention that from the fact that the statement "X is good but I do not desire it" is intelligible, it does not follow that the analysis of "X is good" does not involve desire. The intelligibility of the above locution is only uncontroversial if "I do not desire X" is taken as meaning as it often does "I desire not to have X" rather than "I have no desire for X"; no one denies that one can have conflicting desires. "X is good" might, of course, be an elliptical statement of instrumental goodness, but the relation between ends and means might have to be explained in terms of desire.
  27. Kenny "Action, Emotion and Will" pp.214 - 215
  28. Anscombe "Intention" § 2, § 32. I shall repeatedly refer to this distinction between the two "directions of fit". It may be that I go somewhat beyond Anscombe in the use I make of it. She makes the point in terms of the difference between the types of ground on which we call an order, and an estimate of the future sound, and the difference between the types of mistake to which the execution of an order, or intention and a record of someone's behaviour are liable. The idea that we judge a belief according to whether it matches the world, but judge, or are able to judge the world according to whether it matches our desires and intentions; together with the idea that if we are rational we attempt to bring our beliefs into line with how the world is, whereas we take action to bring the world into line with our desires and intentions (or at least can do so) is an interpretation or extension of Anscombe's point by David Wiggins. In "Freedom, Knowledge, Belief and Certainty" in *Royal Institute of Philosophy Lectures Vol. 3*, p.146, he is arguing that though total causal determinism, though it may well be incompatible with certain aspects of our notion of an agent, does not necessarily undermine our notion of our cognitive orientation to the world. "It must be up to me what to will (but) the libertarian conception of an agent simply cannot make the same demands of causal unconstraint for belief as it makes for desire and the will, unless it fuses belief with fantasy. And this asymmetry in the requirements for freedom reflects the fundamental distraction which Miss Anscombe's (*Intention*) brought into prominence, between thoughts for which the required direction of fit is from the world to the words and thoughts (characteristically couched in the subjunctive, imperative or optative mood) for which the required direction of fit is the

other way about, from the words to the world. Freedom of thought is a rather special freedom. The principal point of making the demand is not to think true whatever you like... but freedom to think or get into a position to think what really is true."

29. Thomas Nagel "The Possibility of Altruism"
30. op.cit. pp. 29 - 30
31. op.cit. pp. 30 - 31
32. op.cit. pp. 36 - 38
33. op.cit. 39 - 40
34. op.cit. pp. 41 - 42
35. op.cit. pp. 43 - 44
36. op.cit. pp. 41 - 42
37. In "Actions, Reasons and Causes" Journal of Philosophy 1963 p.697, Davidson writes: "Any serious theory for predicting action on the basis of reasons must find a way of evaluating the relative force of various desires and beliefs in the matrix of decision; it cannot take as its starting point the refinement of what ~~is~~ to be expected from a single desire. The practical syllogism exhausts its role in displaying an action as falling under a reason; so it cannot be <sup>subtilised</sup> ~~satisfied~~ into a reconstruction of practical reasoning which involves the weighing of competing reasons. The practical syllogism provides a model neither for a predictive science of action nor for a normative account of <sup>evaluative</sup> ~~exhaustive~~ reasoning." It seems that Nagel is demanding that the prudential desire theory do just what Davidson says cannot be done. The prudential desire theory would exhibit prudential reasoning as falling under a certain form of the practical syllogism and this would explain how prudential considerations can motivate, in that it would show how such motivation is intelligible. Its intended application is limited in the same way as the practical syllogism generally, and certainly does not extend as far as Nagel is demanding.
38. op.cit. p.37
39. op.cit. p.69
40. See, for example Davidson "Mental Events" in Foster and Swanson "Experience and Theory".
41. This, of course, raises another problem: that of what we are to say when the bodily movement is overdetermined, when causes of the right kind and of the wrong kind make a contribution to the genesis of the bodily movement and where each set alone would have been sufficient. It does not seem that this is an important difficulty in the analysis, however, as a moment's reflection on the knowledge we would have to have to know that

this situation had arisen in a particular case, shows that it is unlikely that our everyday conceptual scheme surrounding action caters explicitly for this contingency. What is important is that if a causal factor of the wrong kind is found, and this creates a presumption (as it might well do) or a suspicion (as it almost certainly would) that the bodily movement would not have occurred had it not been present, then the attribution of agency is cast in doubt. See also Goldman "A Theory of Human Action", pp. 43 - 44, for the point (in another context) that overdetermination creates a general problem for the attribution of a cause to an event and that derivatives of it will/<sup>be</sup>very likely to arise whenever we attempt an analysis of a concept in causal terms.

42. I am using the term "defeasibility conditions" in such a way that if a defeasibility condition is satisfied, then the ascription of agency fails.
43. It should be clear that we need an account of the "because" in "He did A because of his desire for X" whatever account of deliberation we adopt, and however much we may see our desires as in some way external to ourselves, as something with which we do not necessarily identify. Even if it is held that we can "step back" from our desires, fears, values etc. and decide on which to act, and however extreme an interpretation this is given, the demand for an explication of the "because" will not be circumvented.
44. Anscombe "Intention" § 15
45. Individual cases should, of course, be considered on their own merits. A question that is clearly important for another area of the philosophy of mind is: what is the difference between my being in pain and crying out because I am in pain (i.e. crying out in pain), and my being in pain but crying out for some totally different reason? And what is the difference between these cases and the cases where I cry out, in a sense, because I am in pain, but this crying out is done in order to draw someone's attention to my being in pain? And what is the difference between these simple cases and my crying out for a combination of such reasons?
46. Goldman "A Theory of Human Action" p.67. This definition is laden with Goldman's own particular terminology, and to anyone not familiar with the work, is likely to seem obscure. As my aim is merely to bring out a suggestion implicit in the definition, which I think can be seen independently of the terminology, I shall not burden the account with exegesis.

47. Such possibilities are explored by Thalberg in Chapter 2 of his book "Enigmas of Agency".
48. I am compelled to weaken this claim later on.
49. This is an idea of Danto's in "Basic Actions" American Philosophical Quarterly, 1965.
50. A similar example to Frankfurt's is given by Davidson in "Freedom to Act" (Essays on Freedom of Action - ed. Honderich) p.153.  
"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. It will not help, I think, to add that the belief and want must combine to cause him to want to loosen his hold for there will remain the two questions how the belief and the want caused the second want, and how wanting to loosen his hold caused him to loosen his hold."
51. Nor does it deal with the following case: I desire to raise my left hand. This desire somehow causes a spasmodic jerk of my right hand which drives my left hand up. There might even be cases where the desire to raise my left hand causes a spasmodic jerk of my left hand upwards, and this is not an action. The first difficulty can probably be dealt with by deploying a suitable ~~motiva~~ of basic action. The latter is more on the lines of that under present discussion.
52. See Anscombe "Intention" (passim)
53. In the previous chapter I was critical of the attempt to delineate the concept of action in terms of defeasibility. But there it was in part because I believed a better account to be available. For many of the causal factors whose presence would defeat an ascription of agency on the basis of certain desire we can say that this is because the appropriate desire is not causally involved or not involved as in some way a sufficient condition. Here the desire is causally involved, but not in the right way, and so we cannot avail ourselves of this way of meeting the difficulty. The considerations of the point or direction of fit of desire which I have used to deal with the present difficulty can be used to strengthen the earlier contention that a desire and bodily movement must be causally related if the latter is to be one performed because of the desire.
54. See O'Shaughnessy "Observation and the Will" Journal of Philosophy 1963

55. See Strawson "Review of Wittgenstein's 'Philosophical Investigations'" Mind 1954 p.94
56. Obviously, drumming one's fingers has to be the kind of thing we can do and do intentionally. One does not want to include one's heart beating just because this is an aspect of one's total state at the time.
57. The idea of an action being intentional under some description, but not under others is one way of making sense of the more widely considered class of unintentional actions that Davidson discusses in his paper "Agency" in "Agent, Action and Reason" ed. Binkley, Bronaugh and Marras. It is in fact Davidson's way (though not Goldman's - see "A Theory of Human Action"). It seems to be the only way of making sense of an unintentional action such as turning right when one wants to turn left. Why is this an action at all? The bodily movement that occurs is not that specified by the object of the desire, nor is it an unintended consequence of something one does intentionally, nor is it an intended consequence of something one does. It thus fails to accord with the restrictions we initially laid down to distinguish an action from an emotional reaction, and I do not see how it accommodates itself to Goldman's treatment of unintentional actions. On Davidson's lines we can say that it is intentional under the description "turning", and modify our restrictions accordingly: the bodily movement must meet the specifications of the desire to some extent ( this is vague but it is a beginning). The point is that one wants to call turning right when one wants to turn left an unintentional action. But if one's wanting to turn left causes one's hair to stand on end, one does not see an action here. I have not argued my claim that Goldman's view does not take care of unintentional actions of the above kind, nor tried to meet his criticisms of Davidson's method of act-individuation, as this would take us too far afield, but it is worth referring the reader to Goldman and pointing out that this problem seems to arise here.
58. See Anscombe "Intention" § 52 and Thalberg <sup>Enigmas</sup> "Inquiries of Agency" Chapter 5.
59. Anscombe "Intention" § 2
60. Actually the matter is not quite as simple as this. For it seems that a person could rationally predict that each of two conflicting desires would move him to act. If anything it would be his being moved by two conflicting desires that would be irrational, not the prediction. However, when we speak of

conflicting intentions we are speaking primarily of intentions to perform acts which cannot both be performed. So, if we limit the intentions of which we are speaking to intentions to perform acts, we have the following thesis: a person cannot rationally predict that both of two conflicting desires to act (where one cannot perform both acts) will move him to perform those acts. I will not attempt further refinements because, in the end, I reject the suggested analysis of intention. Nor will I attempt to answer my question: how as a matter of non-contingent fact can the existence of an intention create a presumption that something else, namely a conflicting intention does not exist? The suggestion that this might be partly answered by linking desire with belief shows that this is not an isolated problem; it is not, or not obviously, peculiar to the philosophy of action. For how can a belief have a similar property? It is not difficult to see the broad outlines of an answer: it is contingent, perhaps, in a particular case that a person/<sup>if he</sup>believes p does not believe something that conflicts with it. But if all cases were like this it would undermine the point of belief (to match the world) as both the conflicting beliefs cannot be true; and ultimately the efficacy of desire. There would be no basis for our ascribing beliefs to the agents which we have tried to imagine in this predicament.

61. This point has been mentioned earlier (see footnote 7). I am unhappy about adopting Nagel's view that it is a logical truth that given that someone performs an action intentionally, it is appropriate to ascribe to him a desire to perform that action ("The Possibility of Altruism" pp.29 -30)
62. I do not think that my countenancing this possibility brings me into head-on collision with those who think that predicting one's decisions is impossible for to do so would be to make that decision in advance. To have a belief about what one will do or which desire will move one to act, is not necessarily to be certain. And it is only certainty as to what one will decide to do that seems to entail that one has (contra hypothesis<sup>sm</sup>) already decided. And if it is suggested that for this reason the intuition could not be a certain belief, so at least we have a way of characterising intentions where one is certain about what one is going to do, then the reply is that not all intentions are of this type, and how is one to extend the analysis to intentions that are not? It will not do to say that with these intentions one only really intends to try to do something, for there are intentions where one is not certain what one is



going to do, not because one believes one can only try and might fail, but because the intention is not firm. Finally, it is not obvious that to predict which desire will move one to act is to predict a decision as not all acts are preceded by decisions, and in predicting which desire will move one to act, one does not have to commit oneself to whether a decision will occur. One cannot, it should be noted, solve the problem posed by "intuitions" of one's future behaviour by saying, for example, that such an intuition would not be a real belief because it did not have the appropriate grounds. For then the problem would arise as to how to distinguish such spurious beliefs from intentions. Intentions do have their type of justification, but, as will be seen, this poses problems of its own for the type of analysis in question.

63. Suppose someone suddenly comes to believe that a certain desire will move him to act ( and this belief is non-inductive). But suppose also that he does not believe that it will ever be possible to satisfy this desire. He ought then to believe that at some time in the future he will act in vain perhaps because he will then wrongly believe it possible to satisfy his desire. And surely we are not bound to consider this as adding up to an intention.
64. Anscombe "Intention" § 32. I am claiming that intentions have two directions of fit, not just the one which she ascribes to them.
65. Analysis Problem No.7 "Can I decide to do something immediately without trying to do it immediately" (contributions Brian Ellis, "Candidus" and Nicholas Rescher) Analysis 16, pp.1 - 5.
- 65A If I go to the bank with the intention of cashing a cheque, at what stage do I cease to be merely putting myself in a position to cash the cheque and begin to try to cash the cheque (assuming that there is a reason to speak of "trying" rather than "doing" here)? If the answer seems obvious, then try complicating the question by assuming that before I can cash the cheque I must fill out certain forms.
66. Or by emphasising types of intention for the future which perhaps diverge from the paradigm, such as intentions which are not absolutely firm and intentions to do something some time or other.
67. I am not saying that a case where someone saw something desirable in his having an intention to do A, as distinct from in his doing A, could not be constructed. But if such cases are conceivable, it looks as though one could not form an intention on such a basis. In "Intention" § 27 Miss Anscombe considers

the difficulties of taking an interior act of intention<sup>ding</sup> as evidence of what a man's intentions really are. "I suppose that the man I imagined, who said 'I was only doing my usual job' might find this formula and administer it to himself in the present tense at some stage of his activities. However, if he does this, we notice that the question immediately arises: with what intention does he do it? The question could always arise about anything which was deliberately performed as an 'act of intending'. The answer in this case might be (So that I don't have to consider whose side I am on'. Thus the interior performance has not secured what you might have thought, namely that the man's action in pumping the water is just doing his usual job." Miss Anscombe is perhaps more concerned with the question whether there is a place for an interior act of intention, rather than for an act of intention, but her example does show that a case that might at first sight seem to be one of forming an intention for reasons having to do with the merits of having such an intention, is one in which the genuineness of the intention itself is undermined.

68. Suppose we ask a man why he did not leave the empty milk-bottles out last night, and we receive an answer which non-plusses us, like "To catch pneumonia". Suppose also that the man does have a reason to catch pneumonia, and a reason for believing that not putting out the empty milk-bottles will enable him to achieve this (perhaps they are even good reason). Then if our original request for information had been couched in the form, "With what intention did you refrain from putting out the milk-bottles?" the man would be more justified in believing he had complied with our request, than if we had said, "For what reason did you refrain from putting out the milk-bottles?"
69. There are a number of cases which seemingly do not fit. An example of White's (introduction to "The Philosophy of Action" p.14) is that if I go to Australia with the intention of returning before Christmas, this intention does not give my reason for going to Australia. But if I ask with what intention someone went to Australia, I do not expect an answer of this kind. To ask with what intention something was done seems to ask for the reason for which it was done. And the intention of returning before Christmas is not the intention in virtue of which my going to Australia is intentional. It is the notion of intention which explains an action which concerns us here. White's example could be paraphrased as one of "my going to Australia and intending to return before Christmas."

Miss Anscombe speaks of backward-looking reasons for action. Asked why he killed B, A might reply, "He killed my brother." ("Intention" § 13) This does not mention intention. Rather than attempt to explain this by maintaining that there must be an implicit intention to get revenge, it is sufficient to say that my aim is to show that the intention with which something is done is always the (or a) reason for which it is done. To do this it is not necessary to demonstrate the converse, that the reasons for which something is done is always the intention with which it is done. Finally, there is the case where an intention that something should happen is ascribed to someone, meaning that there was an anticipated and accepted consequence of his action, even though he in no sense desired it. This intention obviously cannot be the reason for his action. I shall deal with this case later in this chapter.

70. This point arose out of discussions with Mr. B. O'Shaughnessy. I have had the following experience which I am sure cannot be uncommon: very annoyed with someone, I resolved to tick them off at the earliest opportunity. When it arrived, my anger had passed, but I obsessively persisted with my intention to upbraid them. But it seemed a charade, a mere play-acting of annoyance, and I was compelled to abandon it half-way through. Surely my reason for attempting to go through with the admonition is to be given in terms of my wanting to be consistent, rather than simply in terms of my original desire. Fidelity to one's intentions in such a case seems to break or at least weaken the connection between the action and the reasons for which the intention to perform it was <sup>formed</sup> found.
71. Even here, it seems difficult to envisage a case where the answer "To signal" could be intelligible if it were thought that the man did not intend to bring anything about by his signalling. One might feel that one needs to assume further intentions which involve a future consequence of the action such as the intention that someone should conclude something from seeing the signal. And even if the answer had been, "To practise signalling", the intention behind practising is normally to develop an ability for the future. Perhaps a better example to show that asking for someone's intention in acting, need not lead to an answer in terms of some envisaged future state of affairs can be found in "Why did you kill him?" to which is answered, "To get revenge", and perhaps a statement of what he did that is resented, a backward-looking reason.
72. To dismiss a usage which does not fit one's theory as a loose one is a move which is rightly treated with suspicion. But I

think the fact that such a usage is itself often treated as imprecise when one is not engaged in defending a philosophical theory is a sufficient justification of the procedure. Miss Anscombe gives a better idiom to characterise the example in question ("Intention" §49) "Something is voluntary though not intentional if it is the antecedently known consequence of one's action, so that one could have prevented it if one would have given up the action: but it is not intentional: ~~one~~ rejects the question 'Why?' in its connection."

73. The use of the term "contingent matter" requires some apology, particularly to anyone who is inclined to read it as entailing that a "contingent matter" is an "empirical matter". I do not wish to assert or deny that one comes to know what bodily movements one can perform in the same way as one comes to know of the occurrence of events in the external world. Nor do I wish to deny that there are conceptual links between the limits of one's body and the actions which ~~one~~ can perform at will (this locution will be clarified later). My claim at this stage amounts to no more than a heuristic device: given that something is a bodily movement, as distinct from a movement in the outside world, it is conceivable that human beings should be able to perform it at will, that is to perform it without doing anything to bring it about (even this is probably too strong for it makes no mention of the question of awareness of the movement - suppose the movement in question were a microscopic brain event. Perhaps my claim will stand if it is restricted to gross, externally visible movements of the body). But given that something is a mental event, say, the onset of a mental state, it is far from obvious that this can be performed at will, that is without doing something to bring it about. My suggestion is that we take account of this latter fact at the <sup>out</sup>onset of our investigation of mental actions. Or to put it another way, let us be sure before attempting to analyse mental actions that we have before us examples of mental actions on which to work, examples that are as uncontroversially of mental actions as raising one's arm is of a physical action. The thesis of the contingency of the limits of our power over our bodily movements is explored by Thalberg in Chapter 2 of "Enigmas of Agency" and my concepts of the limits of one's body and of bodily movements which one can perform without doing anything to bring them about are based in spirit, if not in definition, on Danto's notion of a basic action. See Danto "Basic Actions" American Philosophical Quarterly 1965

74. My debt here is to Rescher's remarks in Analysis Problem No.7 Analysis (1955-56) mentioned earlier.
75. I am not here taking "I decided to stop wanting it" as a conclusion of a piece of practical reasoning; such as a decision that one was wrong to think of something as a means to satisfying one of one's desires. Such a use, if it exists, would assimilate this case to those discussed earlier. I am taking the case of someone who (perhaps after practical reasoning) sees something bad in his having a certain desire but does not thereby lose his desire or his reasons for having the desire.
76. These two cases are, in fact, significantly different as will emerge when we come to discuss "attention". The former is simply hard to conceive, or rather hard to conceive as being done easily. The latter seems to involve conceptual difficulties.
77. Perhaps someone will question my move from "at will" to "if he wants to". And, of course, I have allowed the possibility of reasons for actions other than desires. So, if there is an equivalence here, a full statement of it will be more complicated. My reason for allowing the substitution is that it is a natural reading of "at will", and one that gives a certain plausibility to collapsing "He can bring it about that he desires p (if he wants to)" into "He can desire p at will". But I wish to argue that the latter is an unacceptable idiom as it suggests that desiring p is an action, whereas all we can say is that bringing it about that one desires p might in certain cases be an action.
78. My treatment of this issue might indeed cast doubt on the idea that one can even have a measure of long-range control over one's beliefs, that states subject to such control would not really be beliefs. I shall limit myself to saying that a certain kind of power over one's beliefs tends to deprive belief of its point, and therefore there are conceptual limits to the extent of the control one might have over one's beliefs. I shall leave it an open question as to whether there are many instances of a man bringing it about that he has a certain belief because he wants to have this belief. But it will emerge that any such cases will necessarily have an element of deviousness about them which will often remove them from the sphere of mental action and that a man cannot have full awareness of what he is doing and has done throughout the proceedings. It must not be easy to believe what one wants to believe and in particular it must not be as easy as it is for a normal person, normally

situated to raise his arm.

79. "The Freedom of the Will and the Concept of a Person" Journal of Philosophy (1971) p.9

80. There are certain larger issues in the background here - beliefs and desires which stem from unconscious desires ( a conscious desire that stems from an unconscious desire need not be related, as far as I can see, to the unconscious desire in such normal ways as desire for a means to desire for an end), self-deception and adherence to an ideology or Weltanschauung. I shall not be directly concerned with these, <sup>as</sup> in many cases there is no temptation to think an action has been performed, My strategy is to try to clear some ground by a direct assault on the notion of belief and desire at will, that is closely analogous to <sup>that of</sup> bodily movements at will. It is indeed only a beginning and it is a negative one. But it is hard to see how one can make further progress until one has found some clear-cut cases of mental action, and decisively eliminated some mental events as not mental actions or not involved in them. Another difficulty in considering the relation between beliefs and what one wants to believe can be brought out by considering what I have loosely called ideology or Weltanschauung. It is often difficult to pin-point just what beliefs are involved. It is often suggested that certain beliefs are metaphysical, cognitively meaningless or perhaps just unscientific, because they are unverifiable or unfalsifiable, or because the believer would allow no state of affairs to count against them. We often connect such beliefs with desires to believe. But just what is the believer believing? Furthermore, if in the face of considerations which might seem to cast doubt on a belief, someone reaffirms his belief it is not always obvious when we could say he held the same belief as before. Someone who accepts the account of the Creation in Genesis, faced with the claims of geologists, "modifies" his belief; the account, perhaps, is not literally true, but, given a symbolic interpretation in compatible both with the claims of the geologists and of religion. Has he simply changed from one belief to another, made his belief more precise or reaffirmed his former belief? And how does ~~our~~ answer affect the claim that his desires are influencing his beliefs? The idea of belief as subject to the will in an ideological setting is explored by Price, "Belief and Will" Procedures of the Aristotelian Society Supplementary Vol 28. He mentions attention as a long-range means of securing this, the significance of which will soon emerge.

81. I am assuming that a person cannot desire to desire p if he already desires p. (This is argued by Frankfurt op.cit).<sup>10</sup>) If this assumption is wrong, it does not affect my argument as any counter example to it could hardly be used as a basis for constructing an example of desire at will. I have also ignored the possibility of neurophysiological evidence for what a person's desires are and whether they cause each other. If it is possible that one day we should be able to read off a person's desires from his brain-state, then I am unclear just what to say; except to mention that the fact that if a neurophysiological theory, when supplied, seemed to show that there were cases of desire at will, this might lead us to reconsider the adequacy of the theory. (Suppose such a theory seemed to show the existence of cases of belief at will!)
82. For an account of this effect see D. Wiggins "Freedom, Knowledge, Belief and Causality" in "The Human Agent" Vol. 3 1968/69 pp. 144 - 145 and B.A.O. Williams "Consistency and Realism" in Proceedings of the Aristotelian Society Supplementary Vol.40 (1966) and "Deciding to Believe" in "Problems of the Self; philosophical papers 1956-1972" by B.A.O. Williams, Cambridge University Press 1973.
83. See, for example, Thalberg in Chapter 1 of "Enigmas of Agency" for a full statement of the arguments against this analysis.
84. From the claim that an action is to be analysed in terms of a desire's causing a bodily movement, it is far from clear that anything follows, without further ado, about the legitimacy of the idea of actions being caused. On the one hand, the concept of action is claimed to be a causal concept and actions are situated in the causal nexus. On the other hand, it might be that to speak of an action's being caused, is to employ the notion of "causing to cause", and this, it need hardly be said, is a notion to be handled with care if it is handled at all. Perhaps there is an unexceptionable idea behind this notion; if A occurs and this provides the extra condition which is needed for B to cause C (i.e. B only causes C under certain conditions) we might speak of A causing B to cause C. At least, this does not suggest that A somehow supplies "The causal link" between B and C.
85. It is simply not obvious what causal antecedents of a desire, of the conditions whereby such a desire leads to action and of the properties of the desire which enable it to triumph over others would undermine the status of the action. It does seem difficult to produce a case where, because of our knowledge

of such causal antecedents, we would withhold the term "action" entirely; but no such difficulty to invent cases where the action begins to look "unfree".

86. For reasons which will soon become apparent, it is this second problem that has more hope of an affirmative answer. Someone is faced with the question: given that I am going to do A immediately or almost immediately, can I do it for the right reason, that is, can I, when I act, act on the desire that I want to act on? The answer he must give himself is, I think, "No", unless the question is construed epistemically as, "Might it be the case that the desire I want to move me to act, will alone move me to act". An affirmative answer to this latter question does not mean that it is "up to him" upon which desire he acts. If there is a way in which someone in this predicament can choose to act on the desire on which he wants to act, it will not be revealed by my discussion of the case where someone is faced with the alternative courses of action. It will not help, I think, to suggest that perhaps there is no way in which he does it: he just acts on the desire on which he wants to act. For either this amounts to saying that perhaps when he acts, only the desire he wants to move him does move him, which we have admitted and which leaves his desire to be moved by a certain desire impotent, or it is a mere assertion that such a choice and its implementation is possible. When we ask for a way in which this can happen, we are not asking for a recipe which could be given to the man in question ("Do this and the designed effect will follow.") but whether it is possible that it should happen.
87. Frankfurt op.cit. passim particularly p.8, pp.10 - 111 and 14 - 17 for his definitions.
88. The idea that one has two actions one of which causes the other can be rejected, for what could one possibly do on the basis of the second order desire which caused one to act on the first order desire? If it were a physical act, it would not be within the purview of our present discussion and would amount to a short-range example of the kind of manipulation discussed earlier. And surely there is no mental act that could possibly suffice. There is also a danger in such analyses of treating the agent as an observer of his own actions, if he can manipulate them in such a way. See O'Shaughnessy "Observation and the Will" Journal of Philosophy 1963.
89. Anscombe "Intention" § 14
90. Kenny "Action, Emotion and Will" p.88



91. See footnote 86
92. Anscombe "Intention" § 37
93. I shall refer again to the difficulty of regarding anger, or rather the related notion of resentment, as subject to the will in the next chapter. There I shall assume that resentment, in some cases, is distinguishable from the behaviour to which it might naturally lead such as revenge.
94. The interaction between one's emotions and one's attitude to them is considerable. Not only can one refrain from acting on one's emotions because of countervailing desires, ~~but~~ one can inhibit emotional reactions such as crying because of countervailing desires. On the other hand people often complain of an inability to express their feelings; whereas others seem to have the ability to exploit their emotions for particular ends: the orator who is able to abandon himself to his anger when haranguing his followers is only with a certain artificiality regarded as the pure victim of his emotions, (How is it that he always experiences emotion at the right time?) or simply as a clever actor (his anger will often be an expression of his real opinions).
95. G.N.A. Vesey in "Volition" in "Philosophy" Vol. XXXVI (p.352) writes: "that if a patient who has lost sensation in one arm is asked to put the affected hand on top of his head while his eyes are closed, and is at the same time prevented from doing so, he will be very surprised on opening his eyes to find that the movement has not taken place." This example differs from the simplified case I have produced in that it seems to suggest that anaesthesia might be needed for the phenomenon to occur, and in that the arm fails to move because it is restrained, rather than because it is paralysed. However, the same treatment can be given of Vesey's example as of the one I use. I prefer not to use Vesey's example because, by introducing anaesthesia, it suggests the following treatment: the man in his example does not do anything, but falsely believes he is raising his arm, because he is deprived of his usual knowledge of the position of his limbs and of whether they are being restrained. We would then go on to ask what would happen if there were no such explanation of his false belief, that is if the false belief account were not <sup>extra</sup> given credibility as an alternative to the action account, by the fact that a normal kind of knowledge was suppressed. And we would then be forced into constructing the kind of example that I have used.
96. I have given some thought to the problem as to whether this is

a parallel question to that involved in a consideration of the sense-datum theory: does it follow from the fact that I think I see an X, or seem to see an X, that there is something I see, say a seeming-X, or something I have, say a visual impression of an X? But I have been unable to find any clear-cut analogies or disanalogies between the two questions.

97. Danto "Basic Actions" American Philosophical Quarterly. 1965 p.147
98. Richard Taylor in "Action and Purpose" Chapter 6 argues persuasively against the view that someone whose arm is completely paralysed and who knows it, can try to raise it.
99. In the first chapter I argued against the introduction of the concept of false beliefs that one has a certain desire to account for seeming cases of desire which, for no apparent reason, were not followed by action. It is possible that the case in hand might look similar, and that I am therefore being somewhat inconsistent. But the parallel is not close: for one thing, we are much more ready to admit the idea of someone's being wrong to think he <sup>is performing</sup> a certain action, than to admit the idea of someone's being wrong to think he has a certain desire. And for another, the example in question requires us at the outset to admit that the man is wrong to think he has performed a certain action, raising his arm, because his arm has not gone up. The question is whether this is a complete account of the matter.
100. This reply is somewhat oblique: it is probably wrong to construe a further desire which would be satisfied by performing a certain action, as a reason for acting on a desire which could also lead to that action, unless this means nothing more than that the further desire is a further reason for that action. Such a reply, however, is the sort of reply that might well be given in the face of an interrogation of this nature. I am using this example as a heuristic device to bring out the nature of what constitutes a complete explanation of someone's action in terms of his reasons for performing it.
101. The fact that someone did A rather than B in the kind of circumstances in question creates a presumption that, assuming he deliberated at some time prior to his action, he then preferred to do A? If this presumption is in fact wrong then puzzlement is created (we are close to the old problem of "weakness of will"). In short, reference to prior preferences, felt intensity of desires and the like, do not explain; they tell us that what would normally have been expected to be the case, was the case. If it had not been, then not only might one have

a question as to why the man acted as he did, one has a new question as to why this failed to accord with his prior preferences. If psychologists were able to produce a measure of the strength of desires which could be ascertained without waiting to see which desires proved triumphant, but from a knowledge of which it would be predicted which would prove triumphant, explanations in terms of this concept would not have this quasi-triviality. They would be explanations of why someone acted as he did, but explanations that went beyond simple enumeration of his reasons. If we ignore the loose implications of the statement that someone did A rather than B, both of which he wanted to do, because he preferred A, we have a non-explanatory statement, which is non-explanatory because it exhibits a logical connection between preference and actions. Whereas I have argued against the idea of logical (i.e. entailment) relations between desire and action, such relations might exist between preference and action. Possibly, the impression of vacuity in such explanations of action in terms of preference might be avoided by maintaining that to explain an action in terms of preference is to terminate non-arbitrarily the process of reason giving, to say that one has given all the reasons which weighed with one in deliberating whether to perform that action or some other.

102. I am indebted to Mr. O'Shaughnessy for emphasising to me the frequency with which deliberation already occurs under some heading, when the goal of one's action has already been decided and the problem is to choose<sup>o</sup> between means, a choice which might turn on a matter of fact, or when one has already decided to enjoy oneself and the problem is to decide which of various options one will enjoy most. (I use the expression "the goal of one's action has already been decided" rather than "one has already decided upon the goal of one's action" advisedly, as one may not have even raised the question of acting for some other end; one need not have deliberated to be in this position). Not all deliberation is arbitration between<sup>warring</sup> elements in one's personality. Indeed, it is in the latter cases, that one is unable to give much in the way of reasons for one's choice between alternatives.
103. Or by an explanation in terms of a non-trivial concept of the strength of his desire, or how his desires came to have this strength.
104. Some of these locutions "incredible", "unbelievable" etc. almost suggest that, had the evidence for the man's acting as he did not been so strong, one would have rejected the account on some kind of a priori grounds.

105. It should not be thought that all explanations which might be claimed as satisfying this demand would be genuinely scientific. An astrologer, I suppose, might attempt to provide this sort of explanation.
106. One cannot escape the impression that many writers who oppose the causal theory of action, and are therefore inclined to stress explanations in terms of purpose as the one type of explanation appropriate to action, ignore the possibility of asking why someone did one thing rather than another, when he believed he could not do both but had reasons for each course of action. The burden of my argument is that once one has elicited the purpose or purposes of the action, there might still be residual questions to be asked. The presentation of one's reasons for actions has an end, and if the question as to why certain reasons rather than opposing considerations carried the day is a legitimate one, some other kind of answer will have to be found. Perhaps the proponents of the view in question would be willing to forgo the satisfaction of what I have called the traditional scientific demand - this seems in line with certain equally traditional views on the freewill problem - and concentrate their efforts on an elucidation of the problems about empathy that we have discussed. A good example of a pure form of the view that an explanation in terms of purpose might be all the explanation we can get of an action is found in Richard Taylor's "Action and Purpose". As he does not stress, in addition to purpose, the concepts of convention and rule-following in the explanation of action, as do, for example, Melden and Peters, the implications of this type of view stand out in comparatively clear relief.
107. Strawson "Freedom and Resentment" British Academy Lecture 1962 passim.
108. See Strawson op.cit. p.108
109. If one is to hold someone responsible for not having certain desires or values, one will have to give an account of how he should have acquired them. Perhaps there is some way, in the form of long-range manipulation, although one might not know of it, <sup>still less have it in mind when one feels resentment</sup> ~~But what if he had no desire to have such desires or~~ values? The dangers of a regress are obvious. Or perhaps one assumes that the fact that he does not have certain desires or values can only be the result of his wilfully indulging in or refraining from certain experiences, or that he had the opportunity of acquiring the standards of the community, but somehow culpably refused to do so.
110. This point might form the basis for a less question-begging

formulation of the dilemma posed for the libertarian, going back at least as far as Hume: how are we to distinguish a free act on the libertarian model from a purely random or chance event?

111. Strawson op.cit. particularly pp. 192 - 199
112. Kenny "Action, Emotion and Will" pp. 96 - 97
113. Compare Strawson op.cit., footnote to p.208
114. The difficulty of regarding resentment as subject to the will is similar to that of regarding anger as subject to the will, discussed in the previous chapter and provides one way of arguing for Strawson's claim (op.cit. p.204) that "it is useless to ask whether it would not be rational for us to do what it is not in our nature to (be able to) do." The issue is not quite the same for various reasons. Firstly, we are concerned here with the inhibition of resentment rather than the deliberate adoption of a resentful attitude, with the "turning off" rather than the "turning on" of resentment. Secondly, it is perhaps not so much a matter of refusing to feel resentment because one does not want to, as "<sup>refusing</sup>~~refraining~~" to feel it because of various beliefs. Thirdly, and arising from this, there is the possibility that the reasons which one must have if one is justifiably to feel resentment, might be undermined or defeated by the truth, if this is accepted, of determinism. But it is far from obvious that a general conviction of the truth of determinism would have this effect in a particular case (Strawson argues that our inhibition of our reactive attitudes in particular cases is never the consequence of some such general conviction) It is possible, I suppose, that if human beings become **convinced** of the deterministic theory, they might just stop feeling resentment; that our resentment would, as a matter of fact, be inhibited in this way has yet to be discovered. What is more likely is that people, convinced of determinism, would feel they ought not to feel resentment; and it is unclear what they <sup>could</sup>~~would~~ do about this. It is not unusual to feel resentment and also admit that one's resentment is unjustified. Perhaps one should call it something else, say, "anger"; more importantly, what one can do about the situation seems to be to refrain from acting on one's resentment, rather than somehow refuse to feel resentment. It is of the utmost importance to distinguish the reasons why someone feels resentment from the justification of his resentment. Someone might hit me, and <sup>on</sup>~~that~~ account alone I feel resentment: I also admit that he had a perfectly good excuse, but, against my better judgement, I continue to

feel resentment. The reasons for resentment and the justification for it are, of course, non-contingently related; but the reasons why someone feels resentment are often, even on his own admission, insufficient to amount to a justification. In particular one can be aware of considerations which show one's resentment to be unjustified, and yet find that one still feels resentment. What can one do about one's resentment - or anger, if one is tempted by the terminological move mentioned earlier?

115. The community thereby expressed its condemnation of the crime, and the person who suffered as a result of this atoned for the crime.
116. In order to give ~~sense~~ to the directedness of feelings of resentment and indignation, it is hard to avoid an appeal to the manipulatory practices emphasised by utilitarians. But this then produces difficulties about justice. And here it is common to postulate an "intuition of fittingness"; but the trouble is that this has to be just what it says, "an intuition", for if one attempts to justify its directedness (that it is fitting to punish someone for something he has done) one goes once again round the circle involving utilitarian considerations.
117. Strawson, of course, mitigates the problem by widening the area of practical considerations that are relevant: not only does one need to take account of manipulatory behaviour, but also of such interpersonal behaviour as that involved in love and affection. But this still leaves us with the problem of acts of pure revenge and of acts of retribution on some theories: if these have no further point, why should what the person has done provide a reason for present behaviour, and why should it be directed at him? One can only reply that there is no reason except ~~that~~ the nature of the concept of revenge itself. One is <sup>debarred</sup> ~~deterred~~ from adducing further reasons for the directedness, because the action has, ex hypothes<sup>em</sup>~~is~~, no further point. Moreover, Strawson's considerations about interpersonal relationships do not show the rationale of feeling resentment, as against merely severing one's ties with the person who has wronged one, or trying to get him to change his ways perhaps by a deliberate withdrawal of the signs of goodwill.
118. Strawson op.cit. p.200
119. Reported from Wittgenstein's lectures by Norman Malcolm in his review of "The Philosophical Investigations" Philosophical Review Vol LXIII (1954)

120. Wittgenstein "Philosophical Investigations" § 303
121. See on this Strawson op.cit. p.207
122. This sort of view is odder than my brief characterisation might make it seem. See Joel Feinberg "Supererogation and Rules" Ethics 71 (1961) pp.276 - 288
123. See D. Wiggins "Freedom, Knowledge, Belief and Causality" in "Knowledge and Necessity" Royal Institute of Philosophy Lectures Vol.3.

Bibliography

- Abelson, Raziel "Review of Richard Taylor 'Action and Purpose' " in  
Journal of Philosophy Vol 66 (1969) pp. 178 - 192
- Anscombe G.E.M. "Intention" Oxford: Basil Blackwell 1957
- "Candidus"- see Ellis
- Danto, Arthur "Basic Actions" American Philosophical Quarterly Vol.2  
(1965) pp.141 - 148
- Davidson, Donald "Actions, Reasons and Causes" Journal of Philosophy  
Vol.60 (1963) pp. 685 - 700.  
"Agency" in "Agent, Action and Reason" ed. Binkley, Bronaugh and  
Morris, Oxford: Basil Blackwell 1971  
"Causal Relations" Journal of Philosophy Vol. 64 (1967) pp.691 - 703  
"Freedom to Act" in "Essays on Freedom of Action" ed. T. Honderich,  
London: Routledge and Kegan Paul 1973  
"Mental Events" in "Experience and Theory" ed. I. Foster and  
J. Swanson.London: Duckworth 1970
- Ellis, Brian, "Candidus" and Nicholas Rescher, contributions to  
Analysis Problem No.7, Analysis Vol. 16 (1955-56) pp. 1 - 5. "Can  
I decide to do something immediately without trying to do it  
immediately?" ed. R.B. Braithwaite.
- Feinberg, Joel, "Supererogation and Rules", Ethics Vol.71 (1967) pp.276-288
- Frankfurt, Harry G., "Freedom of the Will and the Concept of a Person"  
Journal of Philosophy Vol. 68 (1971) pp. 5 - 21
- Golberg, Bruce, "Can a Desire be a Cause?" Analysis Vol. 25 (1964 - 65)  
pp. 70 - 72
- Goldman, Alvin J., "A Theory of Human Action" Englewood Cliffs,  
Prentice Hall 1970.
- Hampshire, Stuart, "Thought and Action", London: Chatto and Windus, 1959  
"Freedom of Mind"
- Kenny, Anthony, "Action, Emotion and Will", London, Routledge and  
Kegan Paul 1963.
- Malcolm, Norman, "Wittgenstein's 'Philosophical Investigations' ",  
Philosophical Review Vol. LXIII (1954) pp.530 - 559
- Melden, A.I., "Free Action", London: Routledge and Kegan Paul 1961
- Nagel, Thomas, "The Possibility of Altruism" Oxford: Clarendon Press 1970
- O'Shaughnessy, Brian, "Observation and the Will" Journal of  
Philosophy Vol. 60 (1963) pp.367 - 392.



- Pears, D.F., "Are Reasons for Actions Causes?" in A. Stroll (ed) "Epistemology" New York: <sup>Harper</sup> ~~Naiger~~ and Row 1967, pp. 204 - 228
- "Desires as Causes of Actions" in "The Human Agent" (Royal Institute of Philosophy Lectures Vol. 1(1966 - 67) London: MacMillan 1968, pp.83 - 97
- "Two Problems about Reasons for Actions" in "Agent, Actions and Reason" Binkley, Bronaugh and <sup>Marras</sup> ~~Mauns~~ (ed) Oxford: Basil Blackwell 1971, pp.128 - 153
- Peters, Richard, "The Concept of Motivation" London: Routledge and Kegan Paul, 1958
- Price, H.H., "Belief and Will", Proceedings of the Aristotelian Society, Supplementary Volume (Vol.XXVIII 1954)
- Rescher, Nicholas, see Ellis, Brian.
- Ryle, Gilbert, "The Concept of Mind" London: Hutchinson's University Library, 1949
- Strawson, P.F., "Freedom and Resentment", Annual Philosophical Lecture, Henriette Hert<sup>z</sup> Trust, British Academy, 1962, London O.U.P. 1962
- "Review of Wittgenstein's 'Philosophical Investigations' " Mind Vol 63 1954.
- Taylor, Richard, "Action and Purpose" Englewood Cliffs, Prentice Hall 1966
- Thalberg, Irving, "Enigmas of Agency", London: Allen and Unwin 1972
- Vesey, G.N.A., "Volition", Philosophy Vol 36 (1961) pp. 352 - 365
- White, A.R., (ed) "The Philosophy of Action" Oxford: Oxford University Press, 1968
- Wiggins, David, "Freedom, Knowledge, Belief and Causality" in "Knowledge and Necessity", Royal Institute of Philosophy Lectures, Vol. 3 MacMillan St. Martin's Press, 1970.
- Williams, B.A.O. "Consistency and Realism" in Proceedings of the Aristotelian Society, Supplementary Vol. XL (1966)
- "Deciding to Believe" in "Problems of the Self:philosophical papers, 1956 - 1972", Cambridge University Press 1973
- Wilson, J.R.S., "Emotion and Object", Cambridge,C.U.P. 1972
- Wittgenstein, Ludwig, "Philosophical Investigations" tr. G.E.M. Anscombe, Oxford: Basil Blackwell, 1953