REASONS, MOTIVES, AND CAUSES

by

DAVID ALISTER BROWNE

B.A., The University of British Columbia, 1963
M.A., The University of British Columbia, 1967

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the Department
of
PHILOSOPHY

We accept this thesis as conforming to the
required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April, 1970.
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the Head of my Department or by his representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Philosophy

The University of British Columbia
Vancouver 8, Canada

Date 13 November 1970
ABSTRACT

Introduction

My purpose in writing this thesis is to try to resolve a dispute over what kind of explanation we are giving when we explain an agent's action by giving his reason or reasons for action, or by giving his motive or motives for action. Some philosophers have claimed that such explanations are causal explanations, whereas others have denied this. I shall argue that reason- and motive-explanations are not causal explanations, but constitute an irreducibly different kind of explanation altogether.

Chapter I. Reason-Explanations

In this Chapter I try to make clear what is involved in giving a reason-explanation of an action. I argue that to explain an action by giving the agent's reason or reasons for action is to explain the action in terms of the agent's desires or his desires and information.

Chapter II. Motive-Explanations

In this Chapter I try to make clear what is involved in giving a motive-explanation of an action. I argue that to explain an action by giving the agent's motive or motives for action is to explain the action in terms of the agent's desires and information. Thus the upshot of Chapters I and II is that the question 'What kind of explanation are we giving when we explain an agent's action by giving the agent's
reason(s) or motive(s) for action?' can be re-formulated in a more tractable way, as: 'What kind of explanation are we giving when we explain an agent's action in terms of his desires or his desires and information?'

Chapter III. Desires and Actions

I begin this Chapter by arguing that whether or not explanations of actions in terms of the agent's desires or his desires and information are causal explanations depends on whether or not desires are causes of actions. Many of the arguments designed to show that desires are not causes of actions depend on one or both of two features claimed for the concept of desire. These are specific accounts of (1) the logical connexion that holds between desires and actions, and (2) the descriptions under which specific desires are identifiable. My major aim in this Chapter is to make the nature of these features clear.

Chapter IV. Desires as Causes of Actions (I)

In this Chapter I take up the question 'Are desires causes of actions?', and review some of the arguments and considerations that have been advanced both in favour of answering it in the affirmative and in the negative. I argue that none of these forces us to answer the question in one way or the other.

Chapter V. Desires as Causes of Actions (II)

In this Chapter I give my own answer to the question 'Are desires causes of actions?' I present two distinct arguments to show that they
are not, each of which exploits a different feature of the causal relation. The first argument I present exploits the fact that the causal relation is a contingent relation. I begin this argument by stating a principle that I claim any genuine causal relation must satisfy. In support of this claim, I argue that if this principle is violated, we will be forced to admit that the relation in question is not a contingent relation. And since the causal relation is a contingent relation, any relation that fails to satisfy this principle could not be a causal relation. I then argue that the relation between desires, the conditions under which desires are followed by actions, and actions, fails to satisfy this principle; and that, consequently, these items do not stand in a contingent, and hence could not stand in a causal, relation. The second argument I present to show that desires are not causes of actions begins with the claim that we require empirical evidence to establish the existence of any causal relation. I then go on to argue that we can establish the existence of a relation between desires, certain other conditions, and actions in the absence of any empirical evidence whatsoever; and that, hence, the relation between these items is not a causal relation.

Chapter VI. Concluding Remarks

In my concluding remarks, I draw together the findings of Chapters I to V to yield an answer to the question 'What kind of explanation are we giving when we explain an agent's action in terms of his reason(s) or motive(s) for action? In Chapters I and II, I argued that
to explain an agent's action by giving his reason(s) or motive(s) for action is to explain the action in terms of the agent's desires or his desires and information. In Chapter III I argued that whether or not explanations in terms of the agent's desires or his desires and information are causal explanations depends on whether or not desires are causes of actions. Thus the crucial question to be answered to determine whether or not reason- and motive-explanations are causal explanations turned out to be 'Are desires causes of actions?' In Chapters IV and V, I took up this question, and argued that they are not. And with this finding, we must conclude that reason- and motive-explanations are not causal explanations, but a completely and irreducibly different sort of explanations altogether.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>I REASON-EXPLANATIONS</td>
<td>5</td>
</tr>
<tr>
<td>II MOTIVE-EXPLANATIONS</td>
<td>26</td>
</tr>
<tr>
<td>III DESIRES AND ACTIONS</td>
<td>74</td>
</tr>
<tr>
<td>IV DESIRES AS CAUSES OF ACTIONS (I)</td>
<td>103</td>
</tr>
<tr>
<td>V DESIRES AS CAUSES OF ACTIONS (II)</td>
<td>121</td>
</tr>
<tr>
<td>VI CONCLUDING REMARKS</td>
<td>168</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>172</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENT

I should like to thank Dr. Donald G. Brown and Dr. Warren J. Mullins, who supervised the writing of this thesis, for their encouragement and many helpful comments and criticisms. I have also profited from discussions on specific issues with Dr. Samuel C. Coval, Dr. Howard O. Jackson, Dr. Richard I. Sikora, and Mr. Brian E. Davies.
REASONS, MOTIVES, AND CAUSES

Introduction

We often explain an agent's action by giving his reason or reasons for action, or by giving his motive or motives for action. When we do so, what kind of explanation are we giving of the action? Some philosophers have argued that such explanations are causal explanations, whereas others have denied this.¹ It will be my aim in this thesis to try to resolve this dispute. I shall argue that when we explain an agent's action by giving his reason(s) or motive(s) for action we are not giving a causal explanation of that action, but an irreducibly different sort of explanation altogether. But before this conclusion can be established, a certain amount of preliminary work must be done; and I shall now, both in the interest of making clear what has to be done, and of giving the reader an overview of the structure of the argument as a whole, give a summary outline of the argument of this thesis.

We must first determine what is involved in giving reason- and motive-explanations. In particular, we need to answer the question, 'When we explain an agent's action by giving his reason(s) or motive(s) for action, what are we appealing to to explain the action?' I shall

¹A list of those philosophers who hold or who are sympathetic towards the view that reason- and motive-explanations are causal explanations, and one of those who take or who are sympathetic towards the opposing view, will be found in the Bibliography under the heading 'Reasons and Causes'.

take up this question with respect to reasons in Chapter I; and with respect to motives in Chapter II. I shall argue that when we explain an agent's action by giving his reason or reasons for action we are explaining the action in terms of his desires or his desires and information. And I shall argue that when we explain an agent's action by giving his motive or motives for action we are explaining the action in terms of his desires and information. If these contentions are sound, we may then re-formulate the question 'What kind of explanation are we giving when we explain an agent's action by giving his reason or reasons for action or motive or motives for action?' in a more tractable way, as: 'What kind of explanation are we giving when we explain an agent's action in terms of his desires or his desires and information?'

Now it is obvious, when we explain an agent's action just in terms of some desire of his, that whether or not this is a causal explanation depends on whether or not the desire caused the action. It is not obvious that this is so when we explain an agent's action in terms of his desires and information, but nonetheless I think it is the case, as I shall argue in Chapter III. If so, then the crucial question to be answered to determine whether or not reason- or motive-explanations are causal explanations is 'Are desires causes of actions?' Arguments have been put to support both answers to this question. Many that have been adduced to support the negative answer appeal to one or both of two features that the concept of desire is claimed to have. These concern the logical relation that holds between desires and actions, and the
descriptions under which specific desires are identifiable. It is important, then, for anyone concerned with the question 'Are desires causes of actions?' to make clear the nature of these features appealed to. And in Chapter III I shall try to do this.

I shall then go on in Chapter IV to raise the question 'Are desires causes of actions?' In dealing with this question in Chapter IV, I shall review some of the arguments and considerations that have been advanced both in favour of answering it in the affirmative and in the negative. The arguments in favour of answering the question in the negative that I shall consider are chiefly those of, or ones that derive from or are variations of those of, Ryle and Melden. None of these arguments or considerations will force us to say that desires are, or are not, causes of actions. As I shall argue, those considerations in favour of saying that they are, are indecisive at best; and those arguments in favour of saying that they are not, are unsound.

In Chapter V I shall give my own answer to the question 'Are desires causes of actions?' I shall present two distinct arguments to show that they are not, each of which will exploit a different feature of the causal relation. The first argument I shall present will exploit the fact that the causal relation is a contingent relation. This argument will begin with a statement of a principle that I claim any genuine causal relation must satisfy. In support of this claim, I shall argue that if this principle is violated, we will be forced to admit that the relation in question is not a contingent relation. And since the causal relation is a contingent relation, any relation that fails to satisfy
this principle could not be a causal relation. I shall then argue that
the relation between desires, the conditions under which desires are
followed by actions, and actions, fails to satisfy this principle; and
that, consequently, these items do not stand in a contingent, and hence
could not stand in a causal, relation. The second argument I shall
present to show that desires are not causes of actions will appeal to
the fact that we require empirical evidence to establish the existence
of any causal relation. I shall exploit this fact by arguing that we
can establish the existence of a relation between desires, certain
other conditions, and actions in the absence of any empirical evidence
whatever; and since this is so, that these items do not stand in a
causal relation.

If the arguments of Chapter V are successful, it will emerge that
desires are not causes of actions. This will show that explanations of
actions in terms of the agent's desires alone are not causal explana-
tions. And if my claim in Chapter III, namely that whether or not
explanations of actions in terms of the agent's desires and information
are causal explanations depends on whether or not desires are causes of
actions, is correct, will also show that these explanations are not causal
explanations. Thus, if my arguments to this point are sound, we shall
find that reason- and motive-explanations are not causal explanations, but
explanations of a completely different and irreducible sort. This is, in
outline, the course that my argument will take. I shall now embark on the
detailed analyses and argumentation required to establish the conclusion.
CHAPTER I

REASON-EXPLANATIONS

One of the most common ways of explaining an agent's action is to give his reason or reasons for action. I shall term explanations that explain an agent's action by giving his reason or reasons for action 'reason-explanations'. And in this Chapter, I shall try to determine what is involved in giving a reason-explanation of an action.

There are four features of reason-explanations that I should like to begin by indicating. The first is that they can be used to answer the questions 'Why did he do so-and-so?', 'Why is he doing so-and-so?', and 'Why does he plan to do so-and-so?' That is, they can be used to explain why an agent did something in the past, is doing something in the present, or plans to do something in the future. The second feature is that they can be used to give third-person explanations, i.e., to explain why someone else did, is doing, or plans to do something, and first-person explanations, i.e., to explain why I did, am doing, or plan to do something. The third feature -- one already mentioned in the definition of reason-explanations above, but which I wish to emphasise -- is that the reason that we cite to explain the agent's action is the agent's reason; or, what is the same as this, something the agent regards as a reason for action (or in cases where the action has been performed, regarded as a reason for action). And to say that the agent regards or regarded something as a reason for doing something is quite different from saying that there is or there
was a reason for doing that thing, I shall now try to show that there is a difference, and why the reasons relevant to reason-explanations are the agent's reasons.

There may be a reason for the agent's doing something without him regarding it as a reason for doing that thing. For instance, the fact that I have an appointment at 12:30 that it would be in my interest to keep is a reason for my cutting short my lunch hour. But if I am unaware of this fact, I would not regard it as a reason for action. And if I did cut short my lunch hour, this would not be explained by saying that I had an appointment at 12:30. Conversely, the agent may regard something as a reason for doing something without it being a reason for doing that thing. For example, I may think that I have an appointment at 12:30, regard this as a reason for cutting short my lunch hour, and cut short my lunch hour. But if I am mistaken in thinking this, to say that I have an appointment at 12:30 would not be to give a reason for my action. It would only be to give something I regarded as a reason for acting as I did. But, as such, it could explain my action. What is necessary to convert this 'could explain' into a 'would explain' will be indicated in the fourth feature of reason-explanations below. But at this point, we may say that when we state what actually are reasons for the agent's acting (or having acted or planning to act) in a particular way we are not in all cases giving an explanation of his action. We could only be doing so when the reason co-incided with the agent's reason. And when we explain the agent's action by giving his reason for doing what he did, or is doing, or plans to do, we
are not necessarily (though of course we sometimes will be) giving a reason why the agent ought to have acted, or be acting, or plan to act as he did, is doing, or does. Thus it should be clear that and why the reasons that are relevant to reason-explanations are the agent's reasons.

The fourth feature of reason-explanations is that the reason of the agent's that we cite must be the reason that actually led him to act. It is not enough simply to claim that by giving something the agent regards as a reason for performing a particular action that we are explaining the action. For I may regard something as a reason for doing \( Y \), do \( Y \), and yet not do it for that reason. And in such a case, to cite that thing I regarded as a reason for action would not be to explain my action. In order to explain my action by giving my reason for action, that reason must have been operative. I must have been led to act by that reason; I must have acted for that reason or because of that reason.

Now to say that I did \( Y \) for \( X \) or because of \( X \) (where \( X \) stands for something I regarded as a reason for action) is to say that if I had not regarded \( X \) as a reason for action, I would not have done \( Y \). If I would have done \( Y \) regardless of whether or not I regarded \( X \) as a reason for action, then it would not be true to say that I did \( Y \) for \( X \), that I did \( Y \) because of \( X \). One important point, however, must now be made with respect to this claim that the statement 'I did \( Y \) because of \( X \)' implies the truth of the counterfactual 'If I had not regarded \( X \) as a reason for action, I would not have done \( Y \)'. This is that \( Y \) in the
counterfactual must be taken to stand for the particular action that the agent did in fact perform. The following illustration will perhaps make this point clear. The statement 'I went to the store because I wanted to get some bread' implies only that if I did not want to get some bread, I would not have gone to the store when I did, i.e., would not have performed the particular action I did perform. It does not imply that I would never have gone to the store. It would be as absurd to hold that the statement does have this latter implication as it would be to hold that the statement 'Cancer caused the man's death' implies that if it were not for the cancer, he would never have died.

We have now determined that in order to give a reason-explanation of an action, the reason of the agent's that we cite must be that for which, or because of which, the agent acted, and to say that an agent acted for, or because of, a certain reason is to say that if it were not for that reason, the agent would not have acted in the particular way he did. It follows from these two claims that it is a requirement of any reason-explanation that it yield a true counterfactual statement. For if the counterfactual yielded by a proposed reason-explanation were false, then it would be false that the agent acted for or because of that reason; and if this were false, we would not have a reason-explanation of the action in question.

Now no difficulties arise concerning what must be cited in order to give a reason-explanation of an action when there is only a single factor in question that the agent regards or regarded as a reason for action. But difficulties and puzzles do arise in this respect where
there are two (or more) factors that the agent regards or regarded as reasons for action. So let us now turn to consider what sorts of cases are possible here, and, remembering that any reason-explanation must yield a true counterfactual, try to determine what must be cited in order to give a reason-explanation in such instances.

It may be the case that an agent regards two things as reasons for doing an act, both of which are independently sufficient for his doing the act. Now if he does the act, we cannot without further information say what must be cited in order to give a reason-explanation of his action. Further information about the case may yield two possibilities, each of which must be treated differently. In the first of these, it may be that though an agent regards two things as sufficient reasons for doing an act, and does it, only one is his reason for doing it. For example, a man might have two sufficient reasons for mowing the lawn: to beautify his garden and to please his wife. But even though he perhaps would have mowed the lawn later to beautify his garden, he might do it now to please his wife. In such a case, in order to give a reason-explanation of the action, we would only cite one of his sufficient reasons, and say that he did the deed because he wanted to please his wife. One might object, however, that we could not say this on the ground that to do so is to imply, what is \textit{ex hypothesi} false, that if it were not for his desire to please his wife he would not have mowed the lawn. But it would be wrong to object in this way. For if this counterfactual is taken to mean that if it were not for his desire to please his wife, he would not have mowed the lawn \textit{at all}, though
this is *ex hypothesi* false, it is not implied by the statement that he mowed the lawn because he wanted to please his wife. This latter statement only implies that if it were not for that desire, he would not have mowed the lawn *at the time he did*. And if the counterfactual, viz., if it were not for his desire to please his wife he would not have mowed the lawn, is taken in this latter way, it does not constitute grounds for an objection. For this latter counterfactual is *ex hypothesi* true.

In the case just discussed, though the agent regarded two things as independently sufficient reasons for doing an act, only one of them was operative. But this may not be the case. The agent may have two independently sufficient reasons for doing something that operate concurrently. For example, a man may mow the lawn at a certain time both in order to beautify his garden and to please his wife. In such a case, we cannot cite just one of these in explanation of his acting. For to do this is to imply, what is *ex hypothesi* false, that if it were not for that reason, he would not have acted in the way he did. And since any reason-explanation must yield a true counterfactual, the citing of just one of these concurrent and independently sufficient reasons for action would not be to give a reason-explanation of the action. To give a reason-explanation of the action in such a case, we must cite both reasons of the agent's.

However, when we do this and say that he mowed the lawn (did *Y*) because both he wanted to beautify his garden (*p*) and to please his wife (*q*), this explanation may be misleading. For it is in fact ambiguous. It may be taken to mean that *p* and *q* were concurrent and independently
sufficient reasons for the agent's for doing $Y$, or that $p$ and $q$ were 

independently insufficient but jointly sufficient reasons of the agent's for doing $Y$. On either interpretation, the counterfactual yielded by saying that he did $Y$ because both $p$ and $q$, is, the same: if it were not for $p$ and $q$, the agent would not have done $Y$. And this counterfactual is 

true for both interpretations. Now since the explanation 'He did $Y$ because both $p$ and $q$' is open to these two interpretations, and since the 

counterfactual yielded by this statement, is, true for both interpretations, it seems to me appropriate to use this form of words to explain 

either sort of case. That is, we can explain both cases where $p$ and $q$ 

are concurrent and independently sufficient reasons of the agent's for doing $Y$, and cases where $p$ and $q$ are independently insufficient but 

jointly sufficient reasons of the agent's for doing $Y$, by saying 'He 

did $Y$ because both $p$ and $q$'. If, however, we wish to be more precise, 

and to indicate which of these alternatives we have in mind, it seems to me that we can only do this by a more explicit statement of the 

strength of the agent's reasons. Depending on the case we wish to give 

a reason-explanation of, to say either that the agent did $Y$ for two 

independently sufficient concurrent reasons, $p$ and $q$, or that the agent 
did $Y$ for two independently insufficient but jointly sufficient reasons, 

$p$ and $q$.

In the further discussion of reason-explanations that follows, I shall, for the sake of simplicity, confine my remarks to cases where 

there is only a single reason of the agent's in question. Cases where 

there are two (or more) reasons of the agent's in question are more complex,
and must be handled in the ways indicated above. But apart from this, they introduce no further difficulties.

Let us now turn to consider what is involved in giving a reason-explanation of an action, i.e., to consider what factor or factors we are appealing to when we explain an agent's action by giving his reason or reasons for action. It has sometimes been suggested, for example by Pears, Davidson, and Gean,\(^1\) that when we explain an agent's action by giving his reason for action, this reason, if given in full, would make reference to the agent's desires and information as being the factors that were responsible for the agent's acting in the way he did. The qualification 'if given in full' is necessary here. For the agent often gives his reason for action (or we often give the agent's reason for action) by mentioning only one or the other of these factors. For example, I may say that I am stopping the car because I am hungry; or because there is a restaurant nearby. And sometimes the agent's reason for action is given in another way that mentions neither of these factors, as when I say I am stopping the car to get food. But, according to the suggestion under consideration, when the agent's reason for action is given in these ways, it is given incompletely. The suggestion is that, in cases where only some item of information or some desire is cited, the stated reason for action would be understandable as the agent's

---

reason for action, and would constitute an explanation of his action, only if the other is understood to be present in the situation. And, in cases where neither is cited, as when we give the agent's reason for action by reporting his intention, the suggestion is that we can elucidate this reason for action more concretely and fully in terms of the agent's desires and information. Accordingly, if these suggestions are correct, when we explain an agent's action by giving his reason for action, this reason, if given in full, would make reference to the agent's desires and information as being the factors that were responsible for the action.

Now before raising the question of whether it is true that both desire and information must appear in any complete reason-explanation, I should like to say something about the nature of this information. Much of what I say here ought to be clear from the fact that the reasons involved in giving reason-explanations are the agent's reasons. But I shall risk repetition in the hope of gaining clarity. Often the information in question will be a fact. But to say that it is always a fact is in one way saying too little and in another way saying too much. It would be saying too little in cases where the agent is or was not aware of the fact. For example, if a man were not aware of the fact that there is a restaurant nearby, the citing of this fact would not explain why he stopped the car. This fact indeed, even if unconsidered by the agent, could be a reason for stopping the car, but it could not be his (or a part of his) reason for doing what he did. And it would be saying too much to say that the information involved in reason-explanations is
always a fact in cases where the agent mistakenly thinks something to be a fact. For the agent may act on the basis of this mistaken information; and this (mistaken) information could be used to explain his action by giving his reason or reasons for doing what he did. For example, we could explain why a man was filling his fountain pen with black water by saying that he thought it was ink.

So it seems that the information in question cannot consist exclusively in, or be restricted to, facts. We do, however, often cite facts in giving reason-explanations. But the citing of a fact explains the action only on the assumption that the agent is aware of it. Accordingly, the agent's knowledge must be included in what we mean by information. But since, as we have already seen, the agent's actions can be explained by citing something which he mistakenly thinks to be a fact, we cannot restrict the information in question to knowledge. The agent's beliefs must also be included under this head. Thus we may say that the relevant information in question when we are talking about reason-explanations is an item of knowledge or belief.

Let us now consider the truth of the suggestion that whenever we explain an agent's action by giving his reason for action, this reason, if given in full, would make reference to both desire and information as being the factors that were responsible for his acting in the way he did. We often explain an agent's action by giving his reason for action by mentioning only some item of information, i.e., some item of knowledge or belief. Here are some examples:
(1) I am stopping the car because there is a restaurant nearby.
(2) I am voting because it is my duty to do so.
(3) I handed over the money because he had a gun at my head.
(4) I made myself get to work on time every morning because the company rewards punctuality.
(5) I gave him the book because it was very amusing.

Now it seems that, in all these cases, the citing of this information is intelligible as the agent's reason for action, and constitutes an explanation of the agent's action, only because it is taken for granted that he wants (to do) or wants to avoid (doing) something. For example, saying that there is a restaurant nearby explains my stopping the car only if it is assumed, e.g., that I am hungry. Unless it is assumed or understood that there is something that the agent wants or wants to do which can be satisfied by stopping at the restaurant, to say 'because there is a restaurant nearby' would not explain his action. And if the agent denied that there was anything that he wanted (to do) which could be satisfied by stopping at the restaurant, we should no longer be able to understand this as his reason for action. Similar remarks could be made in the cases of the other examples offered. Thus it seems that to cite some item of information is to give something intelligible as the agent's reason for action, and is to explain the agent's action, only if some desire is understood to be present in the situation. And since some desire must be presupposed whenever one cites only some item of information as the agent's reason for action, to cite only some item of information cannot be to give the agent's reason for
action in full. If we were to give the agent's reason for action in
full, we would have to make explicit this presupposition; and if we did
this, mention would be made of both desire and information.

I shall now turn to consider whether it is true to say that in
all cases where we explain an agent's action by giving his reason for
action by just citing some desire, the citing of the desire is intelli-
gible as the agent's reason for action, and constitutes an explanation
of the agent's action, only if some item of information is understood
to be present in the situation. If it is true, then if we were to give
the agent's reason for action in full, mention would be made of both
desire and information.

We often explain an agent's action by giving his reason for
action by mentioning only some desire. The following are examples:

(1) I am stopping the car because I am hungry.
(2) I am taking my umbrella because I don't want to get wet.
(3) I am voting because I want to do my duty.
(4) I am playing tennis simply because I want to.

Of the first three examples it may be said that I am doing some-
thing not for its own sake but in order to do, achieve, secure, or avoid
something that I want to do, achieve, secure, or avoid. And in the case
of the first three examples, the citing of the desire is intelligible as
the agent's reason for action, and constitutes an explanation of the
agent's action, only on the assumption that some item of information is
present in the situation. For example, to give 'because I am hungry' as
my reason for stopping the car is to explain my action only if some item of information, such as that there is a restaurant nearby, is understood. And if the agent were to deny that he had any information that somehow connected his being hungry and his stopping the car, we should no longer be able to understand 'because I am hungry' as his reason for action. Similar remarks could be made about the second and third examples.

Of course it is often unnecessary to supply the information. For if I say I am taking my umbrella to avoid getting wet, it is quite unnecessary in most cases to add that I believe it is raining or will rain, and that by taking my umbrella I will achieve my desired objective. It would be appropriate for me to state my belief that it will rain only when it would be unusual for me to hold this belief, as for example on a cloudless day, or when it might be thought that I am taking my umbrella to avoid getting wet from some other source. I might, for example, be going to an aqua show where water will be splashed about. But though it is in many cases unnecessary to add this, my point is that some such information must be present in all cases where I am doing what I am not for its own sake, but in order to do or secure or achieve or avoid something that will satisfy some desire of mine; and that the citing of some desire gives an explanation of the action only on the understanding that some such information is present in the situation. Consequently, if we were to give the agent's reason for action in full, this information, in addition to the stated desire, would appear in the statement of it.

The fourth example, of my playing tennis simply because I want to, differs from the first three in that while we can say in the first three
instances that I am doing something in order to so-and-so, in this case I am doing what I am doing for its own sake. I am playing tennis simply because I want to. Now in such cases one may question whether the agent had any reason at all for action; whether such actions are susceptible of reason-explanations. For if an agent is doing something simply for its own sake, and is asked why he is doing it, the reply 'For no reason' seems natural. But this reply cannot be taken to mean that the agent had no reason at all for doing the action. For if he is doing the action simply for its own sake, he must be doing it just because he wants to do it. And if he is doing it just because he wants to, this is his reason for doing it. To say 'For no reason' in this context is, I think, following Davidson's suggestion, only to say that there is no further reason; no reason, that is, besides wanting to do it. So it does seem that when an agent does an action simply for its own sake, he is doing the action for a reason, namely because he wants to.

The question now remains as to whether, when we explain an action in this way, we are giving the agent's reason for action in full, or whether some item of information must be included to do this. In such cases, where we explain an agent's doing X by saying that he is doing so simply because he wants to do X, unlike those already discussed where we explain an agent's doing X by saying that he is doing so because he wants (to do) Y, we can understand the agent's wanting to do X as his reason for doing X without making any assumptions about related

information that he might possess. For here no puzzlement can arise as
to how his desire is related to his action. And so there does not seem
to be any necessity to supplement the agent's desire with some item of
information in order to give a complete statement of his reason for
action. It may, however, be the case that when an agent does an action
for its own sake, say plays tennis simply because he wants to, he be­
lieves that by performing the action he will satisfy his desire to do
so. And in such a case, both desire and information would enter into a
complete statement of the agent's reason for action. We should say that
he is playing tennis because he wants to and knows or believes that by
playing tennis he will satisfy his desire to do so. But it does not
seem to me that some such belief is always consciously entertained when
an agent performs an action for its own sake, and unrealistic to say that
it must be present. Often, it seems to me, an agent does an action with­
out any thought in mind except his wanting to do it. If I am right in
this, it cannot be said that some item of information will always appear
in a complete statement of the agent's reason for action. And thus it
does not seem true in all cases that if we were to give the agent's
reason for action in full, both some desire and some item of related
information would appear in the statement of it. Sometimes when we are
doing something for its own sake, to cite some desire is to give our
complete reason for action: no assumptions about related information
need—or indeed often properly can—be made.

It is important to notice that the word 'want' need not explicitly
occur in reason-explanations that explain by citing some desire of the
agent's. We can use words that imply that the agent wants (to do) or wants to avoid (doing) something. For instance, in the first example offered above, viz., I am stopping the car because I am hungry, the word 'want' does not occur. But to say that I am hungry implies that I want to eat. 'Thirsty' is another word of this type: to say that I am doing $X$, e.g., searching for a water-fountain, because I am thirsty implies that I am doing $X$ because I want to drink. Some other words that can be used in the place of 'want' in giving a reason-explanation are: 'crave', 'lust for', 'fear', 'interested', 'tired', and 'starving'. Examples of sentences in which they are so used are as follows:

1. I stopped at the cafe because I craved (lusted for) a cup of coffee.
2. I left the cave because I fear small enclosed spaces.
3. I went to the lecture because I was interested in hearing the speaker.
4. I stopped running because I was tired.
5. I stole the pie because I was starving.

All the words in question that appear in the above examples are used to imply that the agent wanted (to do) or wanted to avoid (doing) something. Some of them (e.g., 'crave') indicate the degree to which he wanted it; and others (e.g., 'interested') indicate to some extent why he wanted it. And it seems a fairly straightforward matter to rephrase the sentences so that the word 'want' explicitly occurs in them. For example, instead of the above, we could say, respectively:
(1a) I stopped at the cafe because I wanted a cup of coffee.

(2a) I left the cave because I wanted to get away from the small enclosed space.

(3a) I went to the lecture because I wanted to hear the speaker.

(4a) I stopped running because I wanted to rest.

(5a) I stole the pie because I wanted food in the worst way.

It should be noticed, however, that though these words can be, and often are, used to imply that an agent wants (to do) or wants to avoid (doing) something, and so can be and are used in the giving of a reason-explanation of an agent's action that explains the action in terms of the agent's desires, they are not always so used. Examples of cases where they are not used in the giving of the agent's reason for action are easy to find. Take the words 'starving' and 'tired' for instance. To say that I felt weak because I was starving, or that I collapsed because I was tired, is obviously not to give my reason for action; indeed no action occurred. And even when an action does occur that can be explained by using one of these words, they do not always give the agent's reason for acting. For example, I may fidget because I am afraid, but we would not say that fear was my reason for fidgeting.

The words 'love' and 'hate' can also be used in reason-explanations. These words, I think, imply that the agent wants (to do) or wants to avoid (doing) something, but they are slightly more complex than the words listed above. We cannot always—and in most cases cannot—substitute the word 'want' for them as easily as we can do this in the case of the words just considered. For often when these words are used in
giving a reason-explanation, what it is that the agent wants (to do) or wants to avoid (doing) is not specified, but has to be understood from the context. For example, if I say that I withheld the bad news from her because I loved her, what I wanted to do is not specified. But it is clear that I did want to do something: I wanted to do something to benefit her. Similarly with hate. If I give someone the wrong information because I hate him, what I want to do is not specified. But we do know that I wanted to do something to harm him.

The notions of harm and benefit, however, are not always applicable when we explain an agent's action by saying that it was done because he loved or hated him or her. I may ask a girl out because I love her, or ignore a girl at a party because I hate her. And in neither case is it necessarily true to say that I am doing these things to benefit or harm her. But it seems that even in these examples the words 'love' and 'hate' imply that the agent wants (to do) or wants to avoid (doing) something. In the first case, of my asking a girl out because I love her, we know that I want to share her company; and in the second case, of my ignoring a girl because I hate her, we know that I want to avoid her company or conversation.

Besides explaining an agent's action by giving his reason for action by stating some item of information or some desire, or by using some expression that contains a desire-implying word, we often explain an agent's action by giving his reason for action by reporting his intention. Let us now look at how intentions are reported. Intentions are sometimes—perhaps most often—reported by the use of an infinitive
or an infinitive clause. For example, we say that he went to the store to get bread, took a hot bath to relax, married her to get her money, and so on. And in saying 'to get bread', 'to relax', 'to get her money', we are reporting the agent's intention. We do not, however, always report the agent's intention by the use of an infinitive or an infinitive clause. We may say, for example, that he went to the store with the intention of getting some bread, or for the purpose of getting some bread, or for bread. But when we say any of these things, what we say is equivalent to saying that he went to the store to get bread. It seems that we can always give the agent's intention by completing the sentence-frame 'His intention in doing, or planning to do X was or is ...' with an infinitive or an infinitive clause. And for any $\phi$, whenever we say 'His intention in doing, or planning to do X was or is to $\phi$', we may say that his reason for acting was or is to $\phi$. For example, if the agent is taking a hot bath to relax, or selling sub-standard merchandise to increase his profits, we may say that his reason for acting is (respectively) to relax or to increase his profits. And by giving his reason for action by reporting his intention we will be explaining his action.

But while when we explain an agent's action by giving his reason for action by reporting his intention we do not mention any desire nor any information of the agent's, such explanations are not independent of either of these factors. For to know that one is taking a hot bath to relax is to know that he wants to relax and that he knows or believes that by taking a hot bath he will achieve his objective; to know that
an agent is selling sub-standard merchandise to increase his profits is
to know that he wants to increase his profits and that he knows or be­
lieves that by selling sub-standard merchandise he will do so. Or, to
put it generally, to know that an agent is doing $X$ to $\emptyset$ is to know that
he wants to $\emptyset$ and that he believes that by doing $X$ he will $\emptyset$. Accord­
ingly, it seems that whenever an agent's action is explained by giving
his reason for action by reporting his intention, we can elucidate this
reason for action more concretely and fully in terms of the agent's
desires and information. Thus, if we were to give the agent's reason
for action in full, we should cite the agent's desires and information
as the factors that were responsible for the agent's acting in the way
he did.

To sum up the discussion of reason-explanations. We found that
when an action that has been performed, is being performed, or is
planned to be performed not simply for its own sake, but in order to
do or secure or avoid (doing) something, is explained by giving the
agent's reason for action in a way that mentions only some item of
information or some desire, the citing of this factor counts as giving
a reason-explanation of the action only because the other is taken for
granted. We also found, in cases where an action is performed (or has
been, or is planned to be, performed) simply for its own sake, that
sometimes the action is explained by giving the agent's reason for
action by citing only some desire, without our having to make any assump­
tions about related information. Further, we found that we can explain
an agent's action by giving his reason for action by reporting his
intention. And that, though when we so explain an action we do not explicitly mention either any desire nor any information of the agent's, we can always elucidate this reason for action more concretely and fully in terms of the agent's desires and information. Consequently, we may say that when we explain an agent's action by giving his reason for action, this reason, if given in full, will (1) always mention some desire of the agent's, and (2) in every case, except some in which the action is performed for its own sake, mention some information the agent possesses. Thus, when we give a reason-explanation of an action we are often explaining the action in terms of the agent's desires and information, and sometimes in terms of the agent's desires alone.

Accordingly, one of the questions that will be answered in this thesis, namely, 'What kind of explanation are we giving when we explain an agent's action by giving his reason or reasons for action?' can be re-formulated as follows: 'What kind of explanation are we giving when we explain an agent's action in terms of his desires or his desires and information?' One possible answer to this question that I wish to examine is that this type of explanation is a causal one. But having put the question in this form, I shall defer the examination of this possible answer till a later Chapter.
CHAPTER II
MOTIVE-EXPLANATIONS

In this Chapter I shall discuss another common way of explaining an agent's action: that of explaining his action by giving his motive or motives for action. I shall term explanations that explain an agent's action by giving his motive or motives for action 'motive-explanations'. There are two questions that I think must be answered if we are to have a clear account of motive-explanations. These are: (1) What kinds of actions can have motives? and (2) What is involved in giving a motive-explanation of an action?, i.e., precisely to what does the citing of the agent's motive or motives point in order to explain the agent's action? I shall discuss these questions in order.

There is no such thing as simply having a motive; nor do we have motives for feeling irritable or cold, having indigestion or hallucinations, or being pleased or depressed. We talk about motives only when there is an action in question. The action may be one that has been performed, is being performed, or is planned to be performed. That is, we can ask 'What was his motive in doing so-and-so?', 'What is his motive in doing so-and-so?', or 'What is his motive in planning to do so-and-so?' And the action may be one that I performed, am performing, or plan to perform, or one that someone else performed, is performing, or plans to perform.

Let us now raise the first question posed above: What kinds of actions can have motives? It only makes sense to ask for the agent's
motive or motives for acting when the action is, or is thought to be, an intentional one. We do not ask for the agent's motive when we know that the action is not or was not an intentional action. For example, we do not ask for the agent's motive when we know that the action was done by mistake or by accident. Nor do we ask for the agent's motive when we know that the action was a reflex action, where by 'reflex' is meant 'unconditioned reflex'. It is, I think, necessary to insist on this last qualification of reflex actions; for it seems that some actions that are the outcome of conditioned reflexes are intentional actions. For example, a driver who slams on his brakes on seeing a red light seems to be acting, though from a conditioned reflex, intentionally. And similarly, much of what a trained typist does is the result of conditioned reflexes; but we should not deny that her actions were intentional. Whether or not an intentional action of this type, viz., one which is the result of some conditioned reflex, is one that can have a motive will be determined by whether or not it possesses certain other features, to be discussed below, which are necessary for an action to possess in order to have a motive. But we must not preclude the entire class of actions done as a result of conditioned reflexes from the class of actions that can have motives on the ground that they are not intentional actions. At this point we may say that though it makes sense to ask for the agent's motive only when the action is, or is thought to be, an intentional one, only if the action is in fact an intentional one, can it have a motive. An action that is not intentional is an action done without a motive. Let us now try to identify more precisely the class
of actions that can have motives.

The class of intentional actions can be sub-divided, according to the ends that such actions possess or lack, into the four following classes. (1) The action may be performed for its own sake, simply because the agent wants to do it; as for example when one jogs just because he likes or enjoys jogging. (2) The action may be performed with no particular purpose or for the sake of no particular end, i.e., neither for its own sake nor for the sake of anything else. I think that there are many ways in which this can occur; some of them are as follows. The action may be just the exercise of a disposition. For example, a vain man might do vain things intentionally but automatically, and not for the sake of doing those things or in order to achieve anything else. The action may be the outcome or exercise of a mood. For instance, I may speak sharply to a salesman because I am feeling irritable. The action may be one that is the outcome of an emotion, as when I slam the door because I am angry. The action may be one that is simply impulsive and inexplicable to the agent. I may, for example, while strolling down a street suddenly jump and catch a leaf. Or the action may be one that is done purely from force of habit. To say that an agent did something purely from habit is to say that he did not do it for its own sake or in order to do or secure anything else. For to say that an agent did X purely from habit is to say that even if no purpose were served by doing X, he would have done X. And if this is so, the agent did not do X for its own sake or in order to do or secure anything; and this, even if he did in fact do or secure something by doing X.
The above actions I have placed in class (2) are types of intentional actions that I think may be done without any purpose whatsoever. One might, however, deny this, and question the existence of such a class of actions as class (2). He might claim that those actions I have allocated to it will, on closer study, turn out to be either special cases of actions done for their own sakes or actions done with some hidden or unconscious purpose. Such a person would thus empty the class by relocating the actions I have placed in it in class (1) or classes (3) or (4) to be discussed below. I do not know how this issue concerning the existence of class (2) of actions can be clearly decided one way or the other. On the one hand, there is a certain plausibility in the denial that there is such a class of actions as class (2). For many actions that we should not deny are intentional, and which at first sight do not seem to be done with any purpose whatsoever, do turn out on further examination to be done with some hidden or unconscious purpose, or can be plausibly classed as special cases of actions done for their own sakes. And this fact lends weight to the claim that, if we dig deep enough, they will all turn out to be analysable in one or the other of these ways. But on the other hand, it seems dogmatic to insist that all intentional actions that do not seem to be done with any purpose at all must be so analysable. And it seems to me that some simply are not; but, as I said, I do not know how this can be shown to be the case. For it is always open to one who denies that this is so to say, in cases where an intentional action cannot plausibly be claimed to be done for its own sake and where we cannot cite any purpose, hidden or otherwise, of the agent's, that there
is some purpose but that it has just eluded discovery. But anyway, even if I am wrong in asserting that there is such a class of actions as class (2), this will not affect my argument in any substantial way. As will become clear, no point of importance to my argument depends on the existence of class (2) of actions.

To continue with my sub-division of the class of intentional actions. (3) The action may be performed as a means to some further end to be achieved by the action, as for example when a man marries a girl in order to get her money. Marrying the girl is not getting her money, but is a means to this further end. (4) The action may be performed in order to do or secure something that is not a further end. For example, I may be under an obligation to go to a meeting, and go to the meeting in order to fulfill my obligation to do so. In such a case, going to the meeting and fulfilling my obligation to go to the meeting cannot be said to be related as a means to a further end. For in order to say that an action is a means to the doing or securing of some further end, the former and latter must be distinct events; and there are not two distinct events in question here. Rather, there is only one event falling under two descriptions: going to the meeting is fulfilling my obligation to do so. Consequently, if I go to the meeting, though I do so in order to do something, viz., fulfill my obligation to go to the meeting, this latter is not some further end for which going to the meeting is a means.

The above are four sub-classes into which the class of intentional actions can be divided. But the division is not one such that an action
that belongs to one class cannot also belong to another. An action may be done both for its own sake and as a means to some further end. I may, for example, fish both because I enjoy angling and in order to catch a fish. And again, an action may be done for its own sake and also in order to do or secure something that is not a further end. For instance, I may go to a meeting because I like going to meetings as well as in order to discharge an obligation of mine. Thus an action can belong to class (1), viz., the class of actions done for their own sakes, and also belong to class (3), viz., the class of actions done as a means to some further end, or to class (4), viz., the class of actions done in order to do or secure something that is not a further end. But an action cannot belong to both classes (3) and (4). For if I do \( X \) in order to do or secure \( Y \), then either doing \( X \) is a means to the further end of doing or securing \( Y \), or doing \( X \) is doing or securing \( Y \). It cannot be both; for if doing \( X \) (e.g., marrying a girl) is a means to some further end \( Y \) (e.g., securing her money), \( X \) and \( Y \) are necessarily distinct events, but if doing \( X \) (e.g., going to a meeting) is doing or securing \( Y \) (e.g., fulfilling an obligation), then there is only one event falling under two descriptions in question. Nor can actions belonging to class (2), viz., the class of actions done without any purpose whatsoever, belong to any other class. Let us now see which of these classes of intentional actions can have motives.

Actions that just belong to class (1) above cannot have motives. A man who is doing something simply for its own sake is a man not acting from a motive. For example, we would not say that a man who married a
girl simply because he loved her, or committed a murder for its own sake, say just because he felt like it, was a man who acted from a motive. Once we know that the agent is doing or did $X$ simply for its own sake, we also know that he is not doing or did not do $X$ from any motive.

Similarly with actions belonging to class (2). To know that an agent is doing or did $X$ without any purpose whatsoever is to know that he is not acting or did not act from a motive. For example, if $A$ makes disparaging remarks about $B$, and we find out that he made them without any end in view, but that they were just (say) the outcome of some mood of his, we would not say that he had a motive in making the remarks. But actions that belong to classes (3) or (4), i.e., those actions which an agent performs in order to do or secure something (whether this something is a further end or not), can clearly have motives. For example, a man who marries a girl in order to get her money, or kills a man in order to get revenge, is a man acting from a motive. In the first case, his action is performed as a means to a further end; in the second it is not: his killing the man is his being revenged. Such actions may also belong to class (1). In the case of the first example, the man might marry the girl both from this motive and because he loves her. And, in the case of the second, kill the man both from this motive and because he enjoys hurting others. Thus it seems that only actions belonging to classes (3) and (4)—whether or not they also belong to class (1)—can have motives.

If I am right in this, and only actions belonging to classes (3) and (4) above, viz., only actions done in order to do or secure something that is a further end, or done in order to do or secure something that is
not a further end, can have motives, we may make the following two points. First, it is wrong to say, as P.H. Nowell-Smith\(^1\) does, and as R.S. Peters\(^2\) reports many psychologists do, that every intentional action has a motive. For an action done for its own sake is an intentional action, but a motiveless one; and so, it seems, are some actions that are done without any purpose whatsoever intentional but motiveless. And, second, it is wrong to try to correct this error by claiming, as N.S. Sutherland\(^3\) does, that when we explain an action by assigning a motive we are explaining that action in terms of some further end to be achieved by that action. This claim entails that only intentional actions done in order to do or secure something that is a further end can have motives. And it is wrong to claim either of these things. For if I kill a man in order to get revenge, revenge is my motive in killing the man. And in assigning this motive, my action is explained. But the action in question here is neither one explained in terms of, nor one done in order to achieve, some further end. For my killing the man is my being revenged; revenge is not some further end to be achieved by my killing the man.

But is it true to say that every intentional action that is done in order to do or secure something is an action done from a motive? Sutherland holds, or at any rate commits himself to hold, this position.

---


\(^3\) "Motives as Explanations", *Mind*, LXVIII (1959), 148.
For he writes that 'If we say "He did X because he wanted Y" then it is clear that a motive type explanation is being given and there is no necessity to say that "His motive for doing X was that he wanted Y"'.

In saying this, Sutherland commits himself to the position that any substitution instance for Y in the above type of explanation counts as a motive. And since any action that is done in order to do or secure something can be given an explanation of the type 'He did X because he wanted Y', Sutherland also has to maintain that every action of this type is an action done from a motive. But our ordinary usage of the term 'motive' does not support this claim. For example, if I pick up my fork in order to eat my peas, or because I want to eat my peas, it would be odd to say that I am acting from a motive; that my desire to eat my peas was my motive in picking up my fork. Or again, it would be odd to say that when one gives Christmas presents in order to please his friends that this was his motive in acting. And we would normally resent the implication of being asked what our motive is in giving Christmas presents. So it seems that it is not always appropriate to speak of, or ask about, motives when the action is one that is done in order to do or secure something; that not all actions susceptible of 'in order to' explanations are actions done from motives.

Accordingly, we must further delimit the class of actions that can have motives. First of all, it only seems appropriate to ask for the agent's motive when the action in question is a relatively important one.

But there are other features that must be present if the language of motives is to be appropriate. Peters writes that 'we only ask about a man's motives when we wish, in some way, to hold his conduct up for assessment', where 'there is an issue of justification as well as of explanation'. Now this is undoubtedly one of the occasions on which we commonly talk of motives: to ask about an agent's motives is often to imply that his action is a socially unacceptable one, one that stands in need of justification. Hence the resentment we would naturally feel on being asked for our motive in giving Christmas presents.

But it does not seem that Peters is right in claiming that it is only appropriate to ask for a man's motives where there is a question of assessment, let alone a question of assessment involving justification. We can appropriately ask for a man's motives where there is only a question of explanation. For example, we may ask 'What was his motive in giving up his legal practice just when he was becoming successful?' or 'What was his motive in making that odd bequest in his will?' Of course, in certain contexts, to ask these questions may be to hold up the conduct for assessment, and to suggest that some justification is required. But it does not seem to be to do this in all contexts. We can ask these questions when we do not see the point or purpose of the

5 *The Concept of Motivation*, p. 29.

6 Ibid., p. 31.

7 These examples are adapted from R. Brown, *Explanation in Social Science* (London, 1963), p. 91.
agent's action. And in asking for the agent's motive, we may only wish to dispel the mystery that his action holds for us; all we want is an explanation.

We may now take Peters's claim and this latter one together, and say that it is only appropriate to ask for an agent's motive when the action is a relatively important one that stands in need of either justification or explanation. But when are either of these things called for? It seems that justification is called for only when there is some reason to suppose that an action was done for some socially unacceptable reason. In the absence of any reason to suspect this, it would be odd to say that an action requires justification. And it seems that an action stands in need of explanation (but not justification) when there is no reason to suppose that the action was done for some socially unacceptable reason, but when the action is in some way an unusual one. There are, I think, three sorts of actions that we would classify as unusual. The action may be a strange one for anyone to do in the circumstances; the action may be a strange one for the agent, being the sort of person he is, to do in the circumstances; or the action may be one in which the agent's external behaviour is quite normal, but where we have some reason to suppose that it was prompted by some unusual reason. If the agent's action did not fall under one or more of the above three classifications, we would not call it an unusual one; and if it were not unusual in some way, it would not be appropriate to ask for an explanation of the action. With these explanations of when justification and explanation are called for, we can now specify more precisely when it is
appropriate to ask for the agent's motive. It seems that it is only appropriate to ask for the agent's motive when the action is a relatively important one that seems to be done for some socially unacceptable reason, or that is in some way unusual, or, of course, that is both unusual and seems to be done for some socially unacceptable reason.

Now giving the agent's motive may, by rebutting the presumption or suspicion that the action was done for some socially unacceptable reason, justify or excuse the action. Or it may, by confirming the suspicion that the action was done for some socially unacceptable reason, discredit the action. But it may also do something lying between these: it may mitigate the agent's guilt in performing the action. There are also occasions on which giving the agent's motive may be to praise the action, as for example when we find that a priest's motive for leaving his parish was to work in a leper colony. And sometimes there is no question of assessment at all, good or bad, involved in giving the agent's motive. On occasion, giving the agent's motive does no more than explain the action, as when we say that his motive in giving up his legal practice was that he wanted to spend more time with his family.

We have now determined what kinds of actions can have motives. In summary, we have found that two requirements must be satisfied: (1) the action must be an intentional one that is done in order to do or secure something (where this something may be a further end or not), and (2) the action of this kind must be a relatively important one that stands in need of either justification or explanation, or both, i.e., be one that seems to be done for some socially unacceptable reason, or one.
that is in some way unusual, or one that is both unusual and seems to be done for some socially unacceptable reason.

Having thus answered the first question raised at the beginning of this Chapter, I now turn to the second one: What is involved in giving a motive-explanation of an action, i.e., an explanation of an agent's action that explains the action by giving the agent's motive or motives for action? Or, to put this question a bit more concretely, precisely to what does the citing of the agent's motive or motives point in order to explain the agent's action? But before I turn to give my answer to this question, I shall first examine the views of some others, beginning with Gilbert Ryle's.

There is a negative and positive part to Ryle's account. Negatively, Ryle denies that to explain an action by assigning a motive is to explain the action by saying that it was preceded and caused by the occurrence of a certain feeling. Positively, Ryle asserts that to explain an action by assigning a motive is to give a dispositional explanation of the action in terms of 'law-like hypothetical propositions'. Ryle writes that 'To explain an act as done from a certain motive is not analogous to saying that the glass broke, because a stone hit it, but to the quite different type of statement that the glass broke, when the stone hit it, because the glass was brittle'.

Proceeding on this analogy, Ryle recommends construing 'He boasted from vanity' not as saying that 'He boasted and the cause of his boasting was the occurrence in him of a

---

particular feeling or impulse of vanity', but as saying 'He boasted on meeting the stranger and his doing so satisfies the law-like proposition that whenever he finds a chance of securing the admiration and envy of others, he does whatever he thinks will produce this admiration and envy'. According to Ryle, 'To say [that a man did something from a certain motive] is to say that this action, done in its particular circumstances, was just the sort of thing that that was an inclination to do. It is to say "he would do that"'.

Now it seems to me that Ryle is right in what he denies, but wrong in what he asserts. I do not propose to rehearse his arguments for what he denies, but shall only criticise his positive account. It has often, and I think rightly, been pointed out that according to Ryle's account a man cannot act from a motive on one occasion only. This is so, for, according to Ryle, to explain an action by assigning a motive is to explain the action by bringing it under the law-like proposition that the agent is a man who tends to do the sort of thing the motive indicates; it is to say 'he would do that'. But I think that this consequence that a man cannot act from a motive on one occasion only is false. It seems to me that a man can boast from vanity without being a vain man, i.e., without being a man who tends to do vain things, or

---

9 The Concept of Mind, p. 89

10 Ibid. pp. 92-93.

without even being a man who tends to boast. And similarly, that a
man can act from other motives such as generosity or revenge without
being a generous or vengeful man, i.e., without being a man who tends
to act generously or vengefully. If this is so, then Ryle's account
of motive-explanations cannot be correct.

Now we might try to amend Ryle's account to meet this objection
about the man who acts from a motive, for example vanity, on one
occasion by comparing the man's acting from vanity on one occasion to
a piece of glass breaking because of a brittleness which could be
temporary. And then by explaining 'He acted from vanity' as meaning
that at that particular time he tended to react in the ways described
by Ryle, viz., if he finds a chance of securing the admiration and
envy of others, he does whatever he thinks will produce this admiration
and envy. But, as P. Foot has argued, it would be wrong to say this;
because whereas glass which is even temporarily brittle has all the
reactions that go by that name, a man who is temporarily acting from
vanity is not liable to do other things of this kind. For example, a
man who is boasting may not, at that time, be prone to glance admiringly
at himself in a mirror.

Thus Ryle's account seems unsatisfactory. Russell Grice, in his
recent book *The Grounds of Moral Judgement*, has offered an alternative
explanation to that of Ryle's as to how the words 'vanity',

---

12"Free Will as Involving Determinism", in *Free Will and Determin-
ism*, ed. B. Berofsky, pp. 103-104.

13Cambridge, 1967, Ch. I, sec. 3.
'considerateness', 'avarice', 'patriotism', 'indolence', etc., explain an agent's action. I now wish to consider this account. Grice begins by denying, contrary to Ryle, that these words are the names of motives. According to Grice, the only reason that we have for thinking that these terms are the names of motives is that we often answer the question 'Why did he do so-and-so?' by saying 'Vanity', 'Out of patriotism', 'Because he is considerate', etc. He then comments that these answers are in some sense motive-explanations, but that it by no means follows that the motive is explicitly named in them. For the alternative remains that such answers explain why a certain motive, which is not explicitly named, was a motive for a particular man. Grice adopts this alternative view. Taking the example of a man who has acted to ease his neighbour's burden where his action is explained as done out of considerateness, Grice claims that the explanation functions in a two-fold way. First, it explains why the belief that a certain action would ease his neighbour's burden was a motive for this man whereas for many others it would not be a motive. And this it does, according to Grice, by pointing out that the former man is considerate whereas many are not; by pointing out that it is characteristic of him to be moved by such a belief. Second, it explains by revealing the kind of belief which provided the motive for the action. In being told that it was done out of considerateness, we are being told that he was moved by

14 It is perhaps worth drawing attention here to the fact that this is not Grice's account of motives, but only his account of what Ryle takes to be motives. Grice's own account, which I shall not discuss, is similar to the one I shall offer later in this Chapter.
the belief that the action would help his neighbour and not by the belief that he would be judged a benevolent man, nor that his neighbour would return the kindness to him.

This account seems open to the same sort of objection as is Ryle's. Grice's account entails the view that a man cannot act out of considerateness, vanity, patriotism, etc., unless he is a considerate, vain, patriotic, etc., man. For, according to Grice, part of the explanatory force of saying that a man acted out of considerateness, vanity, patriotism, etc., derives from pointing to the fact that he is considerate, vain, patriotic, etc. And Grice could not hold this view unless he also held that a man cannot act out of considerateness, vanity, etc., unless he is a considerate, vain, etc., man. But surely it is wrong to assert that a man cannot act out of considerateness, vanity, etc., unless he is considerate, vain, etc. It seems clear that the most inconsiderate of men can act out of considerateness on occasion. And similarly, men who are generally not vain or patriotic can, on occasion, act out of vanity or patriotism. If this is correct, then to explain an action by using these words cannot be to give a motive-explanation of the action in the way Grice claims it to be.

I shall now turn to give my answer to the question 'What is involved in giving a motive-explanation of an action?' By a motive-explanation I mean an explanation of an agent's action that explains the action by giving his motive or motives for action. The first thing to be noticed is that it is not enough simply to claim that by giving some motive that the agent has for performing a certain action that we are explaining that
action. For I may have a motive for doing $Y$, do $Y$, and yet not do it from that motive. And in such a case, to cite that motive of mine for doing $Y$ would not be to explain my action. In order to explain my action, the motive must have been operative; I must have acted from that motive. So when we ask for a motive-explanation of an action, it seems that, as the word 'motive' suggests, we are asking for what moved the agent to act. To give a motive-explanation of an agent's action is to explain the action by stating what moved him to act. Similarly, in order to explain some action that the agent plans or intends to perform by giving his motive, the motive must be the one that moved him to form that intention. But for the sake of simplifying the exposition, I shall, in what follows, only discuss motives of actions that are being or have been performed.

Now to say that $X$ moved an agent to do $Y$ is to say that if it were not for $X$, the agent would not have done $Y$. If the agent would have done $Y$ even in the absence of $X$, then it would not be true to say that $X$ moved the agent to do $Y$. And here, of course, as in the case of the reason-explanations discussed in Chapter I, '$Y$' in the counterfactual yielded by saying '$X$ moved the agent to do $Y$', viz., 'If it were not for $X$, the agent would not have done $Y$', must be taken to stand for the particular action the agent did in fact perform. For it would be absurd to claim that the statement '$X$ moved the agent to do $Y$' implies that if it were not for $X$, the agent would never have done $Y$ in the same way as it would be absurd to claim that the statement 'Cancer caused the man's death' implies that if it were not for the cancer, he would never have
died. Just as the latter statement only implies that if it were not for the cancer, the man would not have died at the time he did, the former only implies that if it were not for \( X \), the agent would not have done \( Y \) at the time he did, i.e., would not have performed the particular action he did perform.

Once we see that to give a motive-explanation of an agent's action is to explain the action by stating what moved him to act, and that to say that something moved the agent to act is to say that if it were not for that thing he would not have acted in the way he did, it also appears that it is a requirement of any motive-explanation that it imply a true counterfactual statement. For if the counterfactual implied by a proposed motive-explanation were false, then it would be false to say that the thing cited in the proposed motive-explanation moved the agent to act; and if this were false, we would not have a motive-explanation of the action. Nor, indeed, do I think we would have any kind of explanation of the action at all. For it seems that we can always give an explanation of an agent's action by completing the sentence-frame 'He did it because . . . .' And, as we saw in Chapter I, to say that an agent did such-and-such because so-and-so implies the truth of a counterfactual statement. Accordingly, if the implied counterfactual were false, it would be false to say that the agent performed the action because so-and-so; and if this were false, we would not have an explanation of the action at all.

At this point, however, certain difficulties arise. These concern what we must cite in order to give a motive-explanation of an
action in cases where there was more than one factor that moved, or that was sufficient to move, the agent to act. The difficulties that arise here are similar to the ones that arose in a similar respect in my discussion of reason-explanations in Chapter I, and what I say about them will be to a large extent the same.

Let us begin by considering a case in which the agent had two motives for doing an action, each of which was sufficient to move him to act. Now if the agent did the action, we cannot say what must be cited in order to give a motive-explanation of the action without further information about the case. Further information about the case may reveal two possible sorts of cases, each of which must be treated differently. So let us now see what these possibilities are, and how they must be treated.

It may be the case that though the agent had two motives for doing an action, only one of them actually moved him to act. An example of such a case is as follows. A man may have two motives for killing another: to get revenge and to prevent him from giving evidence. And it might be the case that even though he would have killed the man later to get revenge, he did so to prevent him from giving evidence. In such a case we should, in order to give a motive-explanation of the action, cite only one of these motives, and say that he killed the man to prevent him from giving evidence. For it was the desire to do this that moved him to act. And we can say this without violating the idea that to say that something moved an agent to do so-and-so is to imply the truth of a counterfactual statement. For the counterfactual yielded
by saying that the desire to prevent the man's giving evidence moved the agent to kill him is not that if it were not for that motive, he would not have killed the man at all, which is *ex hypothesi* false, but that if it were not for that motive, he would not have killed him when he did. And this, *ex hypothesi*, is true.

I now turn to the second possible case that further information about the case of the man who performs an action for which he had two sufficient motives may reveal. It may be that these motives operated concurrently. For example, a man possessing two motives for killing another (doing \( Y \)), say to get revenge (\( W \)) and to prevent him from giving evidence (\( X \)), each of which was sufficient to move him to do \( Y \), may do \( Y \) from these motives. In such a case what must we cite in order to give a motive-explanation of the action? We cannot simply cite \( W \), his desire to get revenge, nor can we simply cite \( X \), his desire to prevent the man from giving evidence. For to say that \( W \) (or \( X \)) moved the agent to act is to imply, what is *ex hypothesi* false, that if it were not for \( W \) (or \( X \)) the agent would not have done \( Y \). In such a case, we should have to cite both motives of the agent's.

When we do this, however, and say that \( W \) and \( X \) moved the agent to do \( Y \), (or that the agent did \( Y \) because of \( W \) and \( X \)), this explanation is ambiguous. It may be taken to mean that \( W \) and \( X \) were factors that were independently sufficient to move the agent to do \( Y \), and that these factors operated concurrently to do so. Or it may be taken to mean that \( W \) and \( X \) were factors that were independently insufficient, but jointly sufficient, to move the agent to do \( Y \), and that they operated concurrently...
to move the agent to do \( Y \). On either interpretation the counterfactual yielded by saying that \( W \) and \( X \) moved the agent to do \( Y \), (or that the agent did \( Y \) because of \( W \) and \( X \)), is the same: if it were not for \( W \) and \( X \) the agent would not have done \( Y \). And this counterfactual is true for both interpretations. Now since the explanation '\( W \) and \( X \) moved the agent to do \( Y \)' is open to these two interpretations, and since the counterfactual yielded by this statement is true for both interpretations, it seems that it is appropriate to give a motive-explanation in this way in either sort of case. That is, we can give a motive-explanation of the agent's doing \( Y \) from \( W \) and \( X \) both in cases where \( W \) and \( X \) were two concurrently operating factors, each of which was independently sufficient to move the agent to do \( Y \), and in cases where \( W \) and \( X \) were two concurrently operating factors, each of which was independently insufficient and only jointly sufficient to move the agent to do \( Y \), by simply citing \( W \) and \( X \).

However, whereas in the first case, where \( W \) and \( X \) were independently sufficient to move the agent to do \( Y \), we should regard \( W \) and \( X \) as each being a motive of the agent's for doing \( Y \), and together the agent's motives for doing \( Y \), we could not do this in the second case, where \( W \) and \( X \) were independently insufficient and only jointly sufficient to move the agent to do \( Y \). For to say that something is a motive is to say that that thing by itself, in the appropriate circumstances, could have moved the agent to act. And \textit{ex hypothesi} in the second case, neither \( W \) nor \( X \) could independently do this. In such cases as the second one, where \( W \) and \( X \) are independently insufficient and only jointly sufficient to move
the agent to act, the situation would be properly described by saying that \( W \) and \( X \) together constitute a single motive that is internally complex.\(^{15}\)

In the three sorts of cases just discussed, we found that to cite the motive or motives from which the agent acted is to give a motive-explanation of the action. The same would of course also hold true in the simplest case in which the agent only had and acted from a single simple motive. I now want to turn to consider the class of actions, mentioned earlier, where an agent performs an action both for its own sake and in order to do or secure something, with a view to determining whether or not such actions are susceptible of motive-explanations. I shall argue that they are not. There are two possible cases to be considered here. In the first of them, we shall find that an agent can act from a motive and yet that action not be explained by citing that motive; and in the second, we shall find that though it might seem as if an agent has a motive in acting in a particular way, he does not in fact have one. I now turn to consider these cases.

It may be the case that an agent did \( Y \) both in order to do or to secure something, say from some motive \( X \), and for its own sake, where each of these factors was sufficient to move him to do \( Y \). For example, a man might go to a meeting both in order to annoy the chairman and because he likes going to meetings in any event. In such a case as this,

\(^{15}\) The suggestion to so treat this type of case is C.D. Broad's, *Five Types of Ethical Theory* (Totowa, New Jersey, 1965), p. 122.
merely to cite that motive and say that \( X \) moved the agent to do \( Y \) would not be to explain the action. For to say this is to imply, what is \textit{ex hypothesi} false, that if it were not for \( X \) the agent would not have done \( Y \). Consequently, since to assign that motive is not to give an explanation of the action at all, it cannot be to give a motive-explanation of the action. In order to explain the action, we should have to mention, in addition, the other factor or factors that was or were operative in moving the agent to act. But since to assign that motive and to cite the other relevant factor or factors is to explain the action, we may, I think, say that to assign that motive is to give a partial explanation of the action. And since this part of the explanation explains by reference to a motive, we may say that such partial explanations are partial motive-explanations.

The second and final sort of case to be considered can be stated as follows. It may be that an agent performed an action for its own sake and in order to do or to secure something, where neither of these factors was sufficient, but both necessary and jointly sufficient to move the agent to act. That is, it may be that \( Z \) (where \( Z \) is the desire to perform the action for its own sake) and that \( X \) (where \( X \) is the desire for something to be achieved by the action) are neither singly sufficient, but each necessary and jointly sufficient to move the agent to do \( Y \). For example, a man might marry a girl both because he loves her and in order to get her money. And it might be that if he did not love her, his desire to share her money would not have been sufficient to move him to marry her, but that he did not love her so much that he would have...
married her if it were not for his desire to share her money. In such a case, it is tempting, but as I shall argue wrong, to say that his desire to share her money \((X)\) was a motive, or the motive, or a part of an internally complex motive the agent had in marrying her (doing \(Y\)).

It would be wrong to say that \(X\) was a motive the agent had in doing \(Y\), for to say this is to say, what is \textit{ex hypothesi} false, that \(X\) by itself could have moved the agent to do \(Y\). And it would be wrong to say that \(X\) was the agent's motive in doing \(Y\), for though one implication of this, namely that if it were not for \(X\), the agent would not have done \(Y\) is true, it is false to say that \(X\) moved the agent to do \(Y\). An example will perhaps make the falsity of this clearer. Suppose that there is a rock which, in order to move, we must both lift and pull. Exerting upwards pressure on the rock alone will not move it, nor will just pulling at it; but together they will. Now if I exert upwards pressure on the rock while someone else pulls at it, and the rock moves, it would be false to say that I moved the rock or that my partner did. Neither of us moved the rock; together we did. Similarly in the above case, neither \(X\) nor \(Z\) moved the agent to do \(Y\); together \(X\) and \(Z\) did. But it would not be correct to take \(X\) and \(Z\) together as a single internally complex motive, and to say that this was the agent's motive in doing \(Y\). For the desire to do something for its own sake (which is what \(Z\) represents) cannot be a motive at all.

Thus it seems that we cannot refer to \(X\) as being a motive, or the motive, or a part of a motive. But while we cannot assert any of these things, we cannot deny that \(X\) had some influence in moving the agent to
do Y. We are now left with the difficulty of arriving at an adequate characterisation of the influence that X exerted. I suggest that factors such as X, i.e., factors that, were they sufficient to move the agent to act would be classifiable as motives, but which are in fact necessary but not sufficient to move the agent to act, can be termed 'motive-factors'. To say that X is a motive-factor in the agent's doing Y is to accept the implication that if it were not for X, the agent would not have done Y. But it is not to assert that X by itself could have moved the agent to do Y. It is only to assert that X in conjunction with a certain other factor, or certain other factors, could have moved the agent to act. And if the agent did act, X was one of the factors that moved him to do so.

In such a case as this, i.e., in a case where an action is performed both for its own sake and in order to do or to secure something, where neither of these factors is sufficient, but each necessary and jointly sufficient to move the agent to act, it is clear that only to cite what I have termed a 'motive-factor' is not to explain the action. What would be required to explain the action is for us to mention the other factor or factors that contributed towards moving the agent to act. However, since to cite a motive-factor in conjunction with the other relevant factor or factors would be to give an explanation of the action, we may say, as in the above case, that to cite a motive-factor is to give a partial explanation of the action. But we cannot go on to say, as we could and did in the above case, that this is a partial motive-explanation; for motive-factors are not motives.
For the sake of simplicity, I shall, in the subsequent remarks I will make in this Chapter about motives and motive-explanations, presuppose that only one simple motive is in question. Cases where there is more than one motive in question, or where the motive is internally complex, are more complicated, but in principle the same. The remarks I shall make will also apply, with terminological modifications, to partial motive-explanations; but I shall not explicitly mention such explanations again.

Now can we say anything more specific about how motives explain actions than that they explain actions by pointing to what moved the agent to act? I think that we can, and shall argue that to explain an action by giving the agent's motive is to explain the action by reference to the agent's reason for action; and to do this, in cases where the action is one done in order to do or to secure something (which are the only cases of actions relevant here), is, as we found in Chapter I, to explain the action in terms of the agent's desires and information. In saying this, my proposed account will be seen to be partially in agreement with that offered by Ryle. For, as D. Davidson has pointed out, Ryle's analysis of motive-explanations entails that the agent has certain desires and certain related information. Ryle, it will be remembered, analyses 'He boasted from vanity' into 'He boasted on meeting the stranger and his doing so satisfies the law-like proposition that whenever he finds a chance of securing the admiration and envy of others, he does whatever he thinks will produce this admiration and envy'. And Davidson, rightly I think, comments on this that 'if Ryle's boaster did
what he did from vanity, then something entailed by Ryle's analysis is true: the boaster wanted to secure the admiration and envy of others, and he believed that his action would produce this admiration and envy; true or false, Ryle's analysis does not dispense with primary reasons, but depends upon them.\(^{16}\) Thus, to explain an action by assigning a motive is, on Ryle's account, partly to draw attention to the agent's desires and information. But this is not all that there is to it, as we have seen. He goes on to connect explanations in terms of the agent's motive with character traits of the agent. And to do this, as I have argued above, leads to false consequences. My proposed account differs from his in that no reference is made to character traits, but only to the agent's desires and information.

But before I turn to argue for this claim, let us see if motives can be grouped in some way. G.E.M. Anscombe has attempted a grouping of the class of motives; she divides them into three classes: (1) forward-looking, (2) backward-looking, and (3) interpretative motives (or motives-in-general). By forward-looking motives she means those that are equivalent to intentions. By backward-looking motives she means those such as revenge, gratitude, and remorse, which assign something in the past or present as the ground of an action. Anscombe also claims that the motives of this class are characterised by the fact that the notions of good and evil are bound up with them. Now it is unclear from what she says whether the class of backward-looking motives is determined by one

\(^{16}\)"Actions, Reasons, and Causes", in *Free Will and Determinism*, ed. B. Berofsky, p. 226.
criterion, namely that some past or present event is assigned as the
ground of the action, and that it just so happens that the notions of
good and evil seem always to be involved when motives are so assigned.
Or, on the other hand, whether the class of backward-looking motives
is determined by two criteria, namely that some past or present event
is assigned as the ground of the action and that this past or present
event was good or bad for the agent, thus giving rise to the agent's
doing something good or bad in return. I think, (following A. Kenny),
that Anscombe can be most plausibly interpreted as maintaining that what
distinguishes backward-looking motives from the other two kinds is just
that they assign something in the past or present as the ground of the
action. The fact that good and evil are bound up with them is just an
incidental, though according to Anscombe invariable, feature of them
which she uses to distinguish backward-looking motives from mental
causes. And by interpretative motives, (or motives-in-general), Ans-
combe means those motives that place the action in a certain light. As
she puts it, 'To give a motive (of the sort I have labelled "motive-in-
general", as opposed to backward-looking motives and intentions) is to
say something like "See the action in this light"'. Anscombe offers
no definite list of interpretative motives. She suggests that the
motives of curiosity, friendship, admiration, love of truth, fear, spite,
and despair, among a 'host of others', are of this type, but goes on to

---

17 *Action, Emotion and Will*, p. 84.
18 *Intention*, p. 21.
comment that these motives are 'either of this extremely complicated kind [i.e., are interpretative motives] or are forward-looking or mixed'\textsuperscript{19}. And by a 'mixed' motive, as I think is fairly clear from the last sentence, Anscombe means one that is partly interpretative, partly forward-looking. I shall later argue, in opposition to this, that those motives Anscombe tentatively lists as interpretative motives are either forward-looking or partially forward- and partially backward-looking motives.

Let me now draw attention to some defects in this classificatory scheme. One defect is that, according to it, a motive can be both forward-looking (i.e., be equivalent to an intention) and backward-looking (i.e., assign something in the past or present as the ground of the action). Take the motive of revenge for example. If I do $X$ out of revenge, my motive in doing $X$ is to get back at someone for having done something hurtful. And to say this is at once to report my intention ('to get back at someone . . .') and to assign something in the past as the ground for my action ('. . . for having done something hurtful'). Thus revenge must be, on Anscombe's classification of motives, both forward-looking and backward-looking. And similar remarks could be made about the motives of gratitude and remorse.

Another defect is that the class of interpretative motives is not a class that is distinct from the class of motives that are equivalent to intentions. For example, if we take curiosity as an instance of

\textsuperscript{19} Intention, p. 21.
this type of motive, if I do something out of curiosity (say look up an obscure journal or explore a cave), I do so to find out something. If I did not want to find out anything—and this thing may be something definite, as for example to check on a specific reference, or indefinite, as for example to find out what, if anything, is in the cave—I could not be said to be acting out of curiosity. Thus to say that I am acting out of curiosity entails that I am acting to find out something. And since to say that I am acting to find out something is to report my intention, to say that I am acting out of curiosity is equivalent to reporting my intention. Now since Anscombe has not given any definite examples of interpretative motives, and comments after suggesting some that might be interpretative motives that they perhaps are forward-looking or mixed, it could be objected that by showing the motive of curiosity to be equivalent to an intention, I have not thereby shown that the class of interpretative motives is not distinct from the class of forward-looking motives; there may be other motives among those on the list that are not equivalent to intentions. I have merely at this point taken curiosity to be a likely candidate for being an interpretative motive to show that not all those motives tentatively classed under the heading of interpretative motives are distinct from forward-looking motives. I shall examine the other motives on the list later, and argue that they too are equivalent to intentions.

Is the class of interpretative motives distinct from the class of backward-looking motives? Take the example of curiosity again. In one way this can be claimed not to be a backward-looking motive, for to
assign the motive of curiosity is not necessarily to point to something that has happened or is at present happening because of which the action is performed. But in another way, curiosity can function as a backward-looking motive, for what aroused my curiosity might have been something that occurred in the past. A remark, for example, might have made me curious to check on it. And in this case, the ground of my action is something that lies in the past; I am doing something out of curiosity because of some past event.

Since, as we have seen above, Anscombe has not succeeded in distinguishing forward-looking from backward-looking motives by saying that the former are equivalent to intentions while the latter are not, we may try to salvage the distinction by re-formulating the defining characteristic of the class of forward-looking motives. Instead of simply equating forward-looking motives with intentions, we may say that forward-looking motives are those that assign something lying in the future as the ground of the agent's action. And we may oppose this class of motives to the class of backward-looking motives that assign something lying in the past or present as the ground of the agent's action. Drawn in this way, forward-looking and backward-looking motives are distinguished according to where the ground of the agent's action lies. If an agent is acting from a motive in order to achieve some future objective not because of anything that has happened or is at present happening, but just because he wants to achieve that objective, he is acting from a forward-looking motive. But if he is acting from a motive in order to achieve some future objective because of something that happened in the past or is at present happening, he is acting from a backward-looking motive.
By re-defining forward-looking motives in this way, we are now able to say, as we would be unable to say on Anscombe's account, that when one does something solely because of something that happened in the past—as one might in the case of acting out of the motives of revenge or gratitude or remorse—that these motives belong to the class of backward-looking motives and not to the class of forward-looking motives. For we may now admit that backward-looking motives do report the agent's intention, but are not forced on that account to say that they are also forward-looking motives.

I now wish to draw attention to a qualification in the first sentence of the above paragraph. I have written that when one does something solely because of something that happened in the past out of the motives of revenge, gratitude, or remorse, that revenge, gratitude, or remorse are classifiable simply as backward-looking motives. I have qualified my statement in this way because one does not always act out of the motives of revenge, gratitude, or remorse solely because of something in the past, though in the case of these motives there is always some past factor operative. One may look forward to, or anticipate with pleasure, getting revenge, or paying back a past favour, or even, I suppose, repenting for a past wrong. And in such cases there is, in addition to the backward-looking element essential to these three motives, a forward-looking element present in the motive. Thus, though once we know that the agent's motive in acting was revenge (or gratitude or remorse) we can always say that his motive is partially backward-looking, we cannot without further information classify his motive as a
backward-looking motive *simpliciter*.

It is similarly difficult to produce a list of motives that are always forward-looking. Some motives such as ambition usually are forward-looking, for to say that one is acting out of ambition is to say that he is acting to secure a more important position or to succeed or something of this sort. And to say this is normally to indicate some future end that the agent wants to obtain for the sake of which he is acting. But one does not always act from ambition solely to secure the end for its own sake or for some further future end; one may act to secure this end because of something that happened in the past. For example, one might want to obtain a more important position because he was over-ruled too many times in the past. And in such a case, the ground of the agent's action is something that lies, partially at least, in the past. Ambition, then, is not always simply a forward-looking motive. Nor are others, such as patriotism, vanity, and avarice, that are normally forward-looking: a man may act out of patriotism because his country did something for him, or out of vanity because he was often slighted, or out of avarice because he was poor for many years. And so on. And in all these cases we may say that the agent acted, partially at least, in order to redress some imbalance, which imbalance was caused by some past fact, event, or experience. In fact, it seems to me that most actions are prompted by, and done for the sake of, both something that lies in the past and something to be obtained in the future.
Thus we have seen that there is no clear-cut division between forward- and backward-looking motives such that we can produce lists of motives that always fall under just one or the other of these categories. But we can, I think, still draw the distinction, for there are some motives that always have a backward-looking element in them, such as revenge, gratitude, and remorse. And this feature sets them apart from other motives such as ambition, patriotism, vanity, etc., that do not necessarily point to some past fact, event, or experience, but which normally indicate some objective lying in the future. I shall use this distinction to classify motive-words when I come to discuss them.

I now wish to turn and argue for the claim, entered earlier above, that to give a motive-explanation of an action is to explain the action by reference to the agent's reason for action, i.e., in terms of the agent's desires and information. But before embarking on the actual argument for this, it will perhaps be useful to give a sketch of how I shall argue. I shall argue that: (1) to explain an action by giving the agent's motive for action is to explain the action by indicating the agent's end, goal, objective, aim, or purpose in acting; (2) the agent's end, goal, objective, etc., in-acting is given by assigning a motive in only two ways: by reporting his intention or by reporting some desire of his; (3) to explain the agent's action by reporting his intention or by reporting some desire of his is to explain the action by giving the agent's reason for action; (4) to explain the agent's action by giving his reason for action in either of these ways, in cases where the action is done in order to do or to secure something, is to explain the action
In terms of the agent's desires and information; and therefore (5) to give a motive-explanation of an action is to explain the action in terms of the agent's desires and information.

Let me now begin to try to make out these claims by examining the ways in which we give the agent's motive. We often give the agent's motive by filling in a sentence-frame of the sort 'He acted out of . . . ', ' . . . made him do such-and-such', 'His motive in doing such-and-such was . . . ', 'He did such-and-such from . . . ', with a single word such as 'ambition', 'vanity', 'greed', 'jealousy', 'revenge', 'curiosity', etc.

We must now try to determine whether, when we explain an action by giving the agent's motive in any of these ways, we are explaining the action in terms of some end, purpose, objective, aim, or goal towards which the action is directed. I shall argue that we are, and that we do so by reporting the agent's intention.

It is to be noticed, however, that though these words are often used to explain actions by giving the agent's motive, they can be used to explain actions in other ways than by giving the agent's motive. For example, vain people often do vain things automatically, without any end or purpose in mind. Similarly, vengeful persons can act vengefully, greedy persons greedily, avaricious persons avariciously, without acting in order to do or to secure anything. And in such cases we may explain the action by saying that the agent acted out of vanity, revenge, greed, avariciousness. But I do not think that we would call vanity, revenge, greed, etc., in such cases the agent's motive for action. In the above sorts of cases, where there is no purpose or goal present, I think that to
explain the action by saying that the agent acted out of vanity, revenge, etc., amounts to explaining the action as being only the exercise of some disposition. The agent did not act purposively in order to do or to secure something; he only acted out of some disposition as if he were doing so. And once we know that the agent had no end or goal in mind, that he was not acting in order to do or to secure something, we do not ask for, or talk about, the agent's motive.

But while explaining an agent's action as being only the exercise of some disposition does preclude that the agent had some goal in mind, it is quite consistent to say that an agent acted out of a certain disposition and in order to do or to secure something. For clearly a man possessing a vain disposition or character trait may act out of that disposition, and do so in order to secure the admiration and envy of others. And in such a case we can also explain the agent's action by saying that he acted out of vanity. But since, as we saw earlier, it makes sense to say that an action was done out of a motive, e.g., vanity, while yet denying that the agent was a vain man or had a vain disposition, to give a disposition or a character trait of the agent's cannot be included in what is involved in giving the agent's motive. So let us see what it means to say that an agent acted out of vanity, revenge, curiosity, etc., when these phrases are used to give the agent's motive.

I shall begin by considering the meaning of some of the words that are normally used to give forward-looking motives. Among these are: 'ambition', 'greed', 'avarice', 'generosity', 'cowardice', 'patriotism', 'vanity'. To say that an agent acted out of ambition is to say that he
acted in order to obtain a more important position or to succeed or something of this sort. To say that an agent acted out of greed is to say that he acted to try to get more than his share or to get more than was necessary. To say that an agent acted out of avarice is to say that he acted to accumulate money for its own sake. To say that an agent acted out of generosity is to say that he acted to freely share something he had (time, money, etc.). To say that an agent acted out of cowardice is to say that he acted to avoid a dangerous situation that in some sense he ought to have faced. To say that an agent acted out of patriotism is to say that he acted to benefit his country for its own sake. The qualification 'for its own sake' is necessary here, for one may act to benefit his country in order to gain personal glory; and in this case he would not have been acting out of patriotism. To say that an agent acted out of vanity is to say that he acted, usually by doing something trivial, in order to secure the admiration of others or to give himself a chance to contemplate his (real or imagined) good points.

Now if these words mean what I claim them to, we may observe that when they are used to give a motive-explanation of an agent's action, they explain the action by indicating some end, goal, purpose, etc., towards which the action is directed; and indicate this end, goal, purpose, etc., by reporting the agent's intention. For example, to say that an agent did X out of ambition is to say that he acted to secure a more important position. And to say this is to explain the agent's doing X by indicating his goal in acting by reporting his intention.
Similar remarks could be made in the cases of the other words discussed above.

Let us now turn to the class of words which, when they are used to give the agent's motives, always indicate partially backward-looking motives. These are motives that assign something in the past or present as the ground because of which the agent acts. Words that assign motives that are always at least partially backward-looking by virtue of the fact that they always point to some ground in the past are 'revenge', 'gratitude', and 'remorse'. 'Jealousy' is, I think, an example of a motive-word that normally, or at least very often, indicates a backward-looking motive by assigning some ground in the present as that because of which the agent acts. I shall argue that these words, as we found was the case with those that normally indicate forward-looking motives, explain an agent's action by indicating the agent's end, goal, purpose, etc., in acting by reporting his intention. But they do more than just report his intention; they indicate why he had the intention by pointing to some past or present fact, event, or experience. Let us now see what these words mean. To say that an agent acted out of revenge is to say that he acted to pay someone back for having done something hurtful. To say that an agent acted out of gratitude is to say that he acted to confer a benefit because he received one. To say that an agent did something out of remorse is to say that he acted to repent for having committed some wrong. And to say that an agent acted out of jealousy is to say that he acted to harm someone because that person has got or has done something that the agent has not, and would like to have or have
done. In all these cases, to explain an agent's action by using these words is to explain the action by indicating the agent's end, goal, purpose, etc., in acting by reporting his intention, and to allude to some past or present fact, event, or experience as the ground for his having this intention. We do not know what specific fact, event, or experience occurred or is occurring, but we know that some did or is, and its general nature.

I now wish to deal with the words indicating motives falling under Anscombe's third classification: those motives she calls 'interpretative motives' which function to place the action in a certain light. Under this category, she tentatively lists the motives of curiosity, friendship, fear, spite, despair, love of truth, and admiration. The motives of this class seem to be normally forward-looking. There is no backward-looking element essential to these motives as there is in the case of the motives of revenge, gratitude, remorse, and jealousy. But, as was pointed out in the case of curiosity earlier, there may be a backward-looking element present in them which could be discovered if we knew the agent's motive in full enough detail. I now wish to suggest that to give a motive-explanation of an action by using the words 'curiosity', 'friendship', 'fear', 'spite', 'despair', 'admiration', and the phrase 'love of truth', is to explain the action by indicating the agent's end, goal, purpose, etc., in acting; and to indicate this by reporting the agent's intention. An examination of how these words and this phrase explain actions will show that this is so.
To say that an agent acted out of curiosity is to say that he acted to find something out. To say that an agent acted out of friendship is to say that he acted to benefit a friend for his friend's sake. To say that an agent acted out of fear is to say that he acted to avoid or get away from something or someone that he thinks or thought dangerous. To say that an agent acted out of spite is to say that he acted out of a particular kind of ill will to harm or annoy someone, or to affect him in some other negative way. Despair seems only rarely to be a motive for action. And when it is a motive, it is a motive for ceasing to do something. For despair is a state of mind in which the agent is certain that something he wishes to be so is not so. A man may give up a project out of despair, but can never start one out of it. And if we were to explain a man's giving up a project by saying that he did so out of despair, we should be explaining his action by saying that he did so to save himself the effort of pursuing a project that he considered had no chance of success; that he gave it up because he believed he had no chance of success, and wanted to avoid the work or frustration involved in pursuing a task he viewed as hopeless. To say that a man acted out of love of truth, as when we say a man spent his spare time reading non-fiction out of a love of truth, is to say that he acted to find out the truth for its own sake, i.e., not for its utility or for the sake of anything else, but because he had a certain feeling, a pro-attitude, towards truth.

To explain an action by using the word 'admiration', as when we explain one's voting for A by saying that he did so out of admiration or
because he admired $A$, does not seem to fit the pattern exemplified by the other words discussed. That is, we found in the case of 'ambition' that to say that one is doing $X$ out of ambition is to say that he is doing $X$ in order to obtain a more important position; in the case of 'gratitude', we found that to say that one is doing $X$ out of gratitude is to say that he is doing $X$ in order to confer a benefit on someone or something because he has received one; and the other words discussed were found to be susceptible of similar analyses. But when we say that an agent is acting out of admiration, we cannot fill in, as we can in the other cases, what he is acting in order to do: to say that an agent did $X$ out of admiration is to say that he did $X$ in order to . . . what?

It may of course be that the agent did $X$ to show admiration for someone or something. And in such cases the agent's end in acting is clear. But we often act out of admiration without acting to show admiration. And in such cases where the agent acts out of admiration, we cannot ascertain the agent's end or purpose in acting by just explicating the meaning of the word, as we found we could in the cases of an agent's acting out of ambition, patriotism, gratitude, revenge, etc. As with the cases of explanations utilising the words 'love' and 'hate' discussed in Chapter I, when we explain an agent's action by saying that he did it out of admiration, his end in acting has to be gathered from the context. For example, to say that an agent voted for $A$ out of admiration, or because he admired $A$, is to say that he voted for $A$ because he thought $A$ to be an admirable man or an admirable man in some respects. And since to vote for someone is to try to elect him, to know that an agent voted
for A because he admired him is to know that he voted for A in order to try to elect someone he regarded as an admirable man or an admirable man in some respects. Similar remarks could, I think, be made in other cases where an agent's action is explained by saying that he did it out of admiration.

We have now seen that a good many motive-words function to explain the agent's action by indicating his end or purpose or aim or goal or objective in acting by reporting his intention. I suggest, further, that any single word that is used to give the agent's motive functions to explain the action in this way. Now to give the agent's end or purpose etc., in acting by reporting his intention is to give his reason for action. For, for any φ, whenever we say that the agent's end in acting (or in planning to act) is (or was) to φ, we may say that his reason for acting (or in planning to act) is (or was) to φ. Thus to know that the agent's motive was ambition is to know that his end in acting was to obtain a more important position; and to know that his end in acting was to obtain a more important position is to know that this was his reason for acting. Similar remarks could be made in the cases of the other motive-words discussed above. Accordingly, if I am correct in claiming (1) that whenever we explain an agent's action by citing a motive-word, we are explaining the action by indicating the end towards which the action is directed by reporting his intention, and (2) that to give the agent's end in acting by reporting his intention is to give the agent's reason for acting, it follows that whenever we explain an agent's action by giving his motive by the use of a motive-word, we are explaining the
action by giving his reason for action.

We do not, however, always explain actions by giving the agent's motive by completing a sentence-frame of the sort considered above with a motive-word. The agent's motive is often also given in the following ways: He married her in order to get her money, to get her money, for the sake of her money, for the purpose of getting her money, for her money (or for profit), because he wanted her money, because he craved her money. With the possible exception of the phrase 'for the sake of', these italicized expressions never introduce motive-words such as 'ambition', 'vanity', 'jealousy', etc., but must be completed in some such way as the examples indicate. Let us now see how the agent's motive, given in these ways, explains his action. I think it is fairly clear that the above italicized expressions serve to introduce the agent's end, goal, objective, purpose, or aim in acting. We may also notice that the agent's end, goal, objective, purpose, or aim in acting is indicated in one of two ways: by reporting the agent's intention or by reporting some desire of the agent's.

The expressions 'in order to get her money', 'to get her money', 'for the sake of her money', 'for her money', 'for profit', 'for the purpose of getting her money', all explain actions by giving the agent's goal or aim or purpose or objective or end in acting by reporting his intention. All are equivalent to saying 'to get her money'. And, as was argued above, for any φ, whenever we say that the agent's end, purpose, goal, etc., in acting (or in planning to act) is (or was) to φ, we may say that his reason for acting (or in planning to act) is (or was)
to $\phi$. Thus, if we know that the agent's end in marrying the girl was to get her money, we know that his reason for marrying her was to get her money.

The expressions 'because he wanted her money' and 'because he craved her money' explain the agent's action by indicating the agent's end or purpose in acting by reporting some desire of the agent's. And when we say that the agent did (or is doing, or plans to do) something because he desired (or desires) something, by specifying his desire we are giving his reason for action. Accordingly, when we know that he married the girl because he wanted her money, we know that his reason for marrying the girl was his desire for her money.

Thus when we explain an agent's action by giving his motive in any of the ways just considered, we are explaining that action in terms of the agent's reason for action; and doing so by either reporting his intention or reporting some desire of his.

Now it might be objected at this point that I have not dealt with all the types of expressions that can be used to give the agent's motive. It might be urged that, in addition to giving the agent's motive by reporting his intention and by reporting some desire of his, we can give the agent's motive by mentioning some external circumstance or event, or by mentioning some belief of the agent's. That is, it might be thought that the agent's motive can be given by saying things like 'because it was raining', 'because he had a gun', 'because he is a Seventh Day Adventist', and so on. These expressions mentioning some external event or circumstance or some belief of the agent's can be used to explain why the
agent did such-and-such. They can also, I think, be used to make clear the grounds of an agent's motive. But they do not seem to me to explain actions by giving the agent's motive. For we would not say that his motive in handing over the money was the fact that there was a gun at his head, or that his motive in refusing to play tennis on Sunday was the fact that he was a Seventh Day Adventist.

Looking back with a view to summarising our findings, we found that whenever we explain an action by giving the agent's motive, we are always explaining the action by indicating his end, goal, purpose, aim, or objective in acting. We found, further, that the agent's end, goal, purpose, etc., is indicated by giving the agent's motive in only two ways: by reporting his intention or by reporting some desire of his. That is, the agent's motive is always given in, or reducible to, one of the forms 'He did it to φ', or 'He did it out of a desire to φ'. And I argued that to say either of these things is to give the agent's reason for action. Thus to explain an action by giving the agent's motive is to explain the action by giving the agent's reason for action of a special kind, namely a reason of the agent's that indicates the objective or goal aimed at. It is this that moved the agent to act. Accordingly, motive-explanations are a species of reason-explanations.

Now, as was argued in Chapter I, whenever we explain an action by giving the agent's reason for action by reporting his intention, we can always elucidate this reason for action more concretely and fully in terms of the agent's desires and information. For example, to explain my undertaking extra responsibilities by saying that I did so to
obtain a more important position is equivalent to explaining my action by saying that I wanted a more important position, and knew or believed that by undertaking more responsibilities I would increase my chances of, or contribute to, achieving my objective. And, as again was argued in Chapter I, to give the agent's reason for action by reporting some desire of his, where the desire is for some end to be achieved by that action, is to explain the action only if it is understood that the agent has some related information. For example, if I tear up someone's letter because I want to get back at him, my action is explained by saying 'because I want to get back at him' only if it is understood that I believe that by tearing up the letter I will be hurting him.

Thus, since motives are reasons for action, and since they are only given in the two ways just discussed, to explain an action by giving the agent's motive is to explain the action in terms of the agent's desires and information. These are the factors that moved him to act in the way he did, that were responsible for his action.

We can now re-formulate the question 'What kind of explanation are we giving when we explain an agent's action by giving his motive or motives for action?', in a similar way to that which we found in Chapter I we could re-formulate the question 'What kind of explanation are we giving when we explain an agent's action by giving his reason or reasons for action?', as 'What kind of explanation are we giving when we explain an agent's action in terms of his desires and information?' One possible answer to this question is: 'A causal explanation'. But I shall leave the examination of this answer until later; in this Chapter I shall.
remain content with having put the question in this form.

I shall now conclude this Chapter by comparing and contrasting reason-explanations in general with that species of reason-explanations I have termed motive-explanations. Motive-explanations share the following features with other reason-explanations: (1) they can be offered to explain an action done in the past, being done in the present, or planned to be done in the future; (2) they can be used to give third-person explanations, i.e., to explain why someone else did, or is doing, or plans to do something, and first-person explanations, i.e., to explain why I did, or am doing, or plan to do something; (3) the reasons that are relevant to them are those reasons of the agent's that actually moved or led him to act in the way he did; and (4) to explain an action in terms of them is to explain the action in terms of the agent's desires and information. (This last feature is only shared by some reason-explanations, namely those of actions done in order to do or to secure something). But motive-explanations differ from reason-explanations in general in that: (5) they can only be offered to explain actions that are done in order to do or to secure something; and (6) they can only be offered to explain actions that are relatively important and standing in need of justification or explanation (or both). For it is only when an action is a relatively important one that stands in need of one or both of these things do we say that the agent acted from a motive.
CHAPTER III

DESIRES AND ACTIONS

We have now seen the different forms an explanation of an agent's action in terms of his reason(s) for action and motive(s) for action can take. We have seen that we can explain an agent's action by giving his reason for action by reporting some desire of the agent's, or by citing some information the agent possesses, or by reporting his intention. And we have seen that we can explain an agent's action by giving his motive for action by filling in a sentence-frame of the sort 'He acted out of . . .', '. . . made him do such-and-such', 'His motive in doing such-and-such was . . .', 'He did such-and-such from . . .' with a motive-word such as 'ambition', 'gratitude', 'curiosity', or by reporting the agent's desire, or by reporting the agent's intention. We have further found something in common amongst all these types of explanations. In the case of explanations in terms of the agent's reasons, we found that we are explaining the action in terms of the agent's desires or his desires and information. And in the case of explanations in terms of the agent's motives, we found that we are explaining the action in terms of the agent's desires and information. If this analysis is correct, we can re-phrase the question, 'What kind of explanation are we giving when we explain an agent's action by giving his reason(s) or motive(s) for action?' in a more tractable way, as: 'What kind of explanation are we giving when we explain an agent's action in terms of his desires or his
It has often been suggested that this type of explanation is a causal explanation. Now it is clear that when we explain an agent's action just in terms of some desire of his, whether or not this is a causal explanation depends on whether or not desires are causes of actions. I also think that whether or not explanations in terms of the agent's desires and information are causal explanations depends on whether or not desires are causes of actions. I shall now try to show that this is so. In discussing the relation between an agent's information and his action in what follows, I shall be concerned with only one kind of action, namely, intentional action. By making this restriction, I exclude from consideration those possible cases in which the agent's information may be claimed to produce an action in a completely mechanical or conditioned way; for in such cases, the action produced will not be an intentional action.

The agent's information by itself, it seems clear, will never move him to perform an intentional action. The knowledge that a child is starving, or the belief that an operation will save his wife's life, will only be acted on by the agent if he has some relevant desire. If a man were totally indifferent to the welfare of the child or his wife, this information would not be acted on. Put generally, I think it can be truly said that whether or not one acts on the information at his disposal

---

1Throughout this Chapter, and in the following ones, I shall use the words 'desire' and 'want' interchangeably.
depends on whether or not he has some particular desire. This desire may, but need not, be a latent one, antecedent to the awareness of some item of information; it may be one that is produced by the coming into possession of some item of information. But in either case, without the desire there would be no action. Thus if the agent's information is to be said to be causally related to the agent's action, it can only be so when in conjunction with some desire of the agent's.

I shall now proceed to give an account of the ways in which the agent's information can exert an influence on him and his actions. Since I am interested here primarily in determining whether or not, when an action is performed, the agent's information is causally related to that action, I shall ignore the role that information can play in extinguishing certain desires that the agent may have, and the role that the agent's information can play in preventing him from acting in a particular way. The only cases I shall here be considering are those where an agent performs an action because he has certain desires and certain information; and I shall be concerned, with respect to these cases, to give an account of the function of the agent's information. On the account I shall offer, it will emerge that the agent's information can be causally related to his action only if the agent's desires are causally related to his action. And thus, if this account be correct, the crucial question to be answered in order to determine whether or not explanations of an agent's action in terms of his desires and information are causal explanations is 'Are desires causes of action?'
It seems to me, as it did to Hume,\(^2\) that the agent's information functions in only two ways: either by arousing the agent's desire, or by informing the agent of some way in which some desire he has can be satisfied. Let us now examine these two ways in some more detail. Under the first head, where the agent's information arouses some desire, there are two distinct sorts of cases to be considered. The first of these is where the agent has a latent desire to do or to secure something and the information arouses it. I may, for example, have a latent desire to go skiing, and the information that there is fresh snow on the mountains may arouse the desire. Or, to take a slightly different case falling under this same sub-head, I may have a latent desire to keep all my promises, and the information that I promised to do \(X\) may arouse my desire to do \(X\).

The second sort of case to be considered under the general category of cases where information arouses some desire is that in which the agent's information produces a desire where there was none before. For example, it may be the case that I have no pre-existing desire to help people in need. I may hold the view that they ought to help themselves. But striking information about living conditions in depressed areas may cause me to revise my opinion, and produce in me a desire to do something to help.

Now in the cases just considered, if I do what my information arouses in me a desire to do, it seems that even if we grant that the information caused the desire to be aroused, only if the desire caused the action could the information be causally related to the action. That

\(^2\)Hume's views are to be found in his *A Treatise of Human Nature*, Bk. II, Pt. III, Sec. III, and Bk. III, Pt. I, Sec. I.
is, to take an example, even if we grant that the information that there is famine in India caused me to want to send a donation, only if this desire caused me to send a donation could the information be causally related to the action. The causal relation is transitive. If $X$ is the cause of $Y$, and $Y$ is the cause of $Z$, then $X$ is the cause of $Z$. Accordingly, if my information is the cause of my desire, and my desire is the cause of my action, then my information is the cause of my action. But if my desire is not the cause of my action, then, unless there is some other connexion between my information and my action, my information cannot be: if $X$ is the cause of $Y$, but $Y$ is not the cause of $Z$, then, unless there is some other connexion between $X$ and $Z$, $X$ is not the cause of $Z$. We have already seen that there is no causal connexion between the agent's information by itself and his action, and so we may say, in cases where the agent's information arouses his desire (in either of the two ways discussed), that only if the agent's desire can be construed as a cause of his action can his information be said to be causally related to the action.

I now turn to the second way in which the agent's information can function: that of informing the agent of some way in which some desire he has can be satisfied. I may want to catch a fish, and one may tell me that the fish are in the middle of the lake. This information will then make me want to fish in the middle; I will have a desire to fish in the middle of the lake. Or I may have a desire to be in Seattle at 12:00, and find out that the only way of getting there on time is to take the 9:00 train. I will then have a desire to take the 9:00 train. But even
if my information causes me to want to fish in the middle of the lake or to want to catch the 9:00 train, and I do these things, only if the desire caused my action could the information be said to be causally related to my action. For if the information was the cause of my having a particular desire, but the desire did not cause my action, then the information was not causally related to my action.

In many practical situations, the agent's information will function in both this way and in the way previously discussed, i.e., certain information will provide the agent with alternative ways of satisfying some desire he has, and certain other information about these alternative courses of action will arouse in him a desire to perform one of them. This desire may be aroused in either of the two ways earlier discussed: by arousing some latent desire or by producing a desire where there was none before. Schematically, the situation here being envisaged is as follows. I may want X, know that A, B, C are ways of securing X, and certain features of A, B, C may cause me to want to do A over B or C. But even in such a case, whether or not the agent's information is causally related to the action will depend on whether or not his desire is causally related to the action. For if I do A, then only if my desire caused me to do A could the information which caused me to want to do A be said to be causally related to my doing A.

Thus if I am correct in limiting the role of information to the two just discussed, it seems that whether or not explanations of actions in terms of the agent's desires and information are causal explanations depends on whether or not desires are causes of actions.
Now some of the arguments used to try to show that desires are not causes of actions appeal to one or both of two features that the concept of desire is claimed to have. These concern the logical relation that holds between desires and actions, and the descriptions under which specific desires are identifiable. It is important, then, for anyone concerned with the question 'Are desires causes of actions?', to make clear the nature of these features appealed to; and it will be my major aim in the remainder of this Chapter to try to do this. With respect to the first point at issue, the logical relation that holds between desires and actions, I shall provide an account that I will rely on in my argument in Chapter V. But with respect to the second, the descriptions under which specific desires are identifiable, I shall not commit myself to any particular view. Rather, I shall first make clear a widely held view that many arguments designed to show that desires are not causes of actions appeal to, either by itself or in conjunction with the first feature claimed for the concept of desire. I shall then provide two alternative views that have been taken on the question of how specific desires are identifiable. Having done this, I shall defer further discussion of this issue until Chapter V, where what turns on it will be clearer than it will be in this Chapter.

Let me now begin by turning to consider what the logical relationship holding between desires and actions is. I shall argue that we cannot want to do or to secure anything without, other things being equal, trying to do or to secure something in the appropriate circumstances. But before I actually go on to argue for this claim, some preliminary
clarifications are necessary: something must be said about the nature of
the agent in question; the 'other things being equal' clause requires
explanation; and the relevant use of the word 'want' needs to be located.
I shall discuss these points in order.

I shall limit my discussion of wants and actions to those that
can be ascribed to developed, rational, and conscious human agents. And
I shall further presuppose that the agent in question is capable of
trying. The only requirement for being able to try to do something,
where doing that something involves making some bodily movement, is, I
think, that the agent not be totally immobilized (by paralysis, or by
being physically bound, etc.). For to say that one is trying to do some­
thing is always to say that he is actually doing something with a view to
accomplishing a certain result, whether that result is accomplished or
not. The concept of trying is incompatible with doing nothing. Now if
a man were totally immobilized so that he could not move a muscle, then
he could not do anything physically; and hence he could not be said to
be trying to do anything that involves making some bodily movement. But
if he were only partially immobilized, say only his arm was paralyzed,
then it seems to me that he could try to move it: he could, for example,
tense certain muscles or grit his teeth in the attempt to move his arm.
These things would count as trying to do something. However, if he did
none of these sorts of things, then we would say that he is not trying.

I shall now turn to explain the 'other things being equal' clause
that occurs in my claim that we cannot want to do or to secure anything
without, other things being equal, trying to do or to secure something
in the appropriate circumstances. The qualification 'other things being equal' is meant to exclude the presence of countervailing wants and the lack of belief on the part of the agent that he has either or both the ability and/or opportunity to do or to secure what he wants. Thus, filling in the 'other things being equal' clause, my claim amounts to saying that if the agent has a want, no countervailing wants, and believes that he has the ability and opportunity to realise or satisfy his want, then he will try to do so. I am claiming that the relationship between the propositions constituting the antecedent and the consequent of this conditional is that of entailment. This means, if we let 'p' stand for the proposition that is the antecedent and 'q' stand for the proposition that is the consequent of the above conditional, that it is inconsistent to assert p and yet deny q. It should be noticed here that it is not necessary that the agent actually have the ability and opportunity; it is only necessary that he believe himself to have the ability and the opportunity in order for it to be necessary that, given he has a want and no countervailing wants, he try to do something. Some comments on countervailing wants, ability, and opportunity would now be in order. In explaining what is included under these three heads, the import of my claim will, I hope, become clearer.

Countervailing wants may be directed to things or actions or states of affairs that the agent wants or wants to do or wants to bring about more than he wants to satisfy the want in question. They may also be constituted by an aversion to something involved in satisfying or realising the want in question. For example, I may want to have X, but to have X is
illegal; and I may want to avoid breaking the law more than I want $X$. I am using the term 'ability' here to cover such things as skills as well as other things such as money, time, strength, patience, etc. We must now determine what it means to say that someone has the opportunity to do or to secure something. It is clearly the case that when one is in a position to do something that certainly or probably will achieve some end, we can say that he has the opportunity to achieve that end. But we do not restrict the application of the term 'opportunity' to cases such as these, where the agent is in a position to do something that will certainly or probably succeed in achieving some end. For example, we quite naturally say that every candidate at the convention had the opportunity to become leader of the party. And when we say this we certainly do not mean that they were all in a position to certainly or probably become leader. Or again, we can say that a man had the opportunity to stop the thief. And we can quite properly say this even if it is not certain or probable that he would have stopped the thief if he had tried. In both these cases, all that seems to be required for us to say that the individuals had opportunities is that it would have been possible for them, respectively, to attain the leadership or to stop the thief. Moreover, it seems proper to say that one has the opportunity to do or to secure something where he is so circumstanced that it is possible but improbable for him to succeed. For example, in the second case, where a man is said to have had the opportunity to stop the thief, even if we knew that he was so circumstanced that it was improbable that he could have, as long as it was possible, we would say that he had the opportunity to do so.
We might speak of his having not a very good opportunity, but an opportunity nonetheless. Thus it seems that all that is required for someone to have the opportunity to do or to secure something is that he be in a position that makes it possible for him to do or to secure that thing. And, without further information, this is all that it means to say that an agent has the opportunity to do or to secure something. We must now formulate what it means to say that the agent believes that he has the opportunity to do or to secure something; for this is what is relevant to the discussion of the 'other things being equal' clause. If, as I have argued, to say that an agent has the opportunity to do or to secure something is to say that he is in a position that makes it possible for him to do or to secure that thing, to say that an agent believes he has the opportunity to do or to secure something is to say that he believes he is in a position that makes it possible for him to do or to secure that thing.

I now wish to consider a possible objection arising out of the account I have offered of what it means to say that one has the opportunity to do or to secure something. I have claimed that if an agent has a want, no countervailing wants, and believes that he has the ability and the opportunity to satisfy his want, then he will try to do so. I have also claimed that the relation between the propositions constituting the antecedent and consequent of this conditional is entailment. Not it may be argued that since we can say that an agent has the opportunity when the probability of success is low, or when it is possible but improbable for him to succeed, the agent may believe that he has the opportunity but not
act because of the lowness of the probability of success. Thus the
force of the objection is to suggest that an agent can have a want, no
countervailing wants, believe that he has the ability and opportunity,
and yet not act because the probability of success attaching to his
opportunity is not high enough; and that, hence, the relationship between
the antecedent and consequent of this conditional could not be one of
tailment.

I propose to deal with this objection in the following way. If
the probability of success is very low, or where success is possible
but improbable, I may decide not to try to do anything; but I decide
not to try to do anything because I do not want to expend the effort
trying to do something that has so little a chance of success. That is,
it is a stronger counter-want that accounts for my not performing. So
my answer to the objection under consideration that holds that an agent
can have a want, no countervailing wants, believe that he has the ability
and opportunity, and yet not act because the probability of success
attached to the opportunity is not high enough, is to deny that when the
probability is not high enough that the agent has no countervailing wants.

With these explanations, I hope it is clear what I mean by claiming
that we cannot want to do or to secure anything without, other things
being equal, trying to do or to secure something in the appropriate cir-
cumstances.

Something must now be said about the relevant use of the word
'want'. The word 'want' may be used in many distinct ways, not all of
which are relevant to the explanation of actions. We may begin to try to
isolate the relevant use of the word by excluding the following uses as being irrelevant to the explanation of actions: (1) Commands. E.g., 'I want you to be in my office at ten sharp'. (2) Requests. E.g., 'I want a piece of pie, please'. (3) Choices. E.g., 'I want the one on the left'. (4) 'Want' in the sense of 'ought'. E.g., 'You want to eat before you go'.

R.B. Brandt and J. Kim also exclude the following uses of 'want' as being irrelevant to the understanding of actions: (5) 'Want' in the sense of 'need'. E.g., 'The child wants to be disciplined'. (6) 'Want' in the ways in which the word is used in the sentences: 'The police want him', 'You are wanted by the boss'. As the word is used in these sentences, 'want' is used in the sense of 'seek'; or perhaps, in the second sentence, in the sense of 'need'.

But though the word 'want' when used in the sense of 'need' or 'seek' is sometimes, as in the above examples, irrelevant to the explanation of actions, we cannot exclude these uses as being always irrelevant. Sometimes they are clearly relevant to explaining actions, as in the following cases. If I went to the library because I needed to check on a reference, my action could be explained by saying that I went to the library because I wanted to check on a reference. Similarly, if I made some outrageous remark because I sought publicity, my action could be explained by saying that I made that outrageous remark because I

---

wanted publicity. Thus these uses of 'want', viz., in the senses of 'need' and 'seek', cannot be totally excluded from consideration.

The word 'want' can be used in a wider or narrower way. When used in a narrower way, it can be contrasted with words or expressions like 'needs' or 'has to'. For example, we may say that he went to the library not because he wanted to, but because he had (or needed) to. But using it in the wider way, to say that he went to the library because he had (or needed) to entails that he went there because he wanted to. It is this wider use of 'want' that is relevant to the explaining of actions. But what is this wider use? I think that it can be indicated with sufficient precision by saying that it is the use of 'want' in which that term can always be substituted for the phrase 'in order to' (together with, of course, changes in the surrounding sentence-structure). It seems that whenever we say that one did $X$ in order to do or to secure $Y$, we may re-phrase this by saying that one did $X$ because he wanted to do or to secure $Y$. Now the doing or securing of $Y$ may or may not be something that the agent would do as a matter of choice or inclination, i.e., would do if there were no other considerations involved. For example, I may go to a meeting in order to read a paper, but I may hate reading papers. But even in such a case, there is a sense in which it can be properly said that I went to the meeting because I wanted to read a paper. Or, on the other hand, I may hike to a mountain lake in order to get in some good fishing, which I love doing. And here too we can clearly replace 'in order to' with 'because I want to'. Of course, the 'in order to' construction is not always relevant to the explanation of
actions; it is not relevant in those cases where the agent performs an action for its own sake, simply because he wants to. But this does not affect the point I am making, which is merely that the use of 'want' relevant to the explanation of actions is the widest possible one. It includes the use of 'want' that is conformable with, as well as the use that is in contrast to, something that is a matter of the agent's pleasure. The context is to be relied upon to make it clear which is the case.

As this is the use of 'want' relevant to the explanation of actions, it is also the use that is to be taken to be exemplified by that word in my claim that we cannot want to do or to secure anything without, other things being equal, trying to do or to secure something in the appropriate circumstances. Accordingly, when one says 'I want to go to the library', he may or may not be anticipating, with pleasure, the trip. He may want to go there because he has to for some purpose or other, or he may want to go there because he likes or enjoys going to libraries.

Having now given some explanations of what I mean by claiming that we cannot want to do or to secure anything without, other things being equal, trying to do or to secure something in the appropriate circumstances, and having isolated the relevant use of the word 'want', I now wish to turn to argue for this claim. I shall begin by examining what it means for an agent to say that he wants to do or to secure something. Take, for example, the sentence 'I want to go to the circus'. This sentence can be used in the appropriate circumstances, (i.e., not when I am learning a language, illustrating a philosophical point, etc.) to
make a statement that is either true or false: it is either true or false that I want to go to the circus. When I make the statement 'I want to go to the circus', I am describing something about myself in the way in which a piece of glass can be described as brittle. When we say that the glass is brittle, this entails that if it is acted upon in certain ways it will break. If the glass did not break, we should say that the glass is not brittle, and that the statement that claimed it to be so is false. In a similar manner, when I say that I want to go to the circus, this entails that if there is a circus in town, then, other things being equal, I will intend to go to it. If I do not, and all the conditions are met, (i.e., I have no stronger counter-wants, and believe I have the ability and opportunity), then I could not properly say that I want to go. If such a claim were made, it would be false; just as the claim that the glass is brittle would be false if the glass did not break when struck in the appropriate way. And, as I shall shortly show, we often have behavioural criteria by which we can verify or falsify a 'want'-statement.

If my argument so far has been correct, we cannot want to do or to secure anything without, other things being equal, intending to do or to secure something in the appropriate circumstances. Thus wants involve conditional intentions. It is important to notice here that what we intend to do need not be what we want to do. When we want something we may act directly to try to secure it, in which case the object of our intention and the object of our want will co-incide; or we may act indirectly by doing something which is believed to be a means to securing
what we want, in which case the object of our intention and the object of our want, though they will be related, will not co-incide.

Now to say that one intends to do or to secure something entails that he will, other things being equal, try to do or to secure that thing in the appropriate circumstances. If an agent professed to have an intention to do something, had no stronger counter-wants, believed that he had the ability and opportunity to do it, and yet did not act on the intention, he could not properly be said to have it. From such a failure to try to do something under these conditions, we should conclude either that he had changed his mind or was lying when he claimed to have the intention. If I am correct in this, we cannot intend to do or to secure anything without, other things being equal, trying to do or to secure it in the appropriate circumstances.

I am now in a position to show the connexion between wanting and trying to do or trying to secure something. Since (1) we cannot want to do or to secure anything without, other things being equal, intending to do or to secure something in the appropriate circumstances, and since (2) we cannot intend to do or to secure anything without, other things being equal, trying to do or to secure it in the appropriate circumstances, it follows that (3) we cannot want to do or to secure anything without, other things being equal, trying to do or to secure something in the appropriate circumstances. Now if we fill in the 'other things being equal' clause in (3), we can arrive at a statement citing certain conditions that will entail a statement to the effect that we will try to do something. This would be as follows. The statement that we want to
do or to secure something, have no countervailing wants; and believe that we have the ability and opportunity to do or to secure it, entails the statement that we will try to do or to secure something in the appropriate circumstances.

Now I claim that the conditions listed in the 'other things being equal' clause, (viz., no countervailing wants, and the belief on the part of the agent that he has both the ability and opportunity to do or to secure something that would lead to the satisfaction or the realisation of his want), are exhaustive in the sense that, given that the agent wants (to do) something, any specific alternative to the agent's trying to do or to secure something will fall under one of these three heads. That is, if a 'want'-statement is made, then at least one of the following disjuncts will obtain: the agent will try to do or to secure something, or there will be some countervailing wants, or the agent will not believe that he has the ability, or the agent will not believe that he has the opportunity. If none of these alternatives is the case, then the 'want'-statement could not be said to be true at the time when action is appropriate. The phrase 'at the time when action is appropriate' is important, for this allows for the fact that, if there is a time-lag between the making of the statement and the time for action, the agent may change his mind. If the agent does not try to do or to secure anything, and none of the other disjuncts obtains, my claim is not that the 'want'-statement was never true; it is just that the 'want'-statement is not true when the time
for action has come. 4

I now wish to turn to the ways in which we can verify or falsify a 'want'-statement. Very often there will be behavioural criteria by the presence or lack of which we can verify or falsify a 'want'-statement. But this verification or falsification is not a simple matter, and I should now like to discuss the difficulties involved. There are three distinguishable sorts of cases that have to be dealt with here. In dealing with them, I shall formulate the 'want'-statement in the first person, but what I say about them equally applies to 'want'-statements in any other person.

Case (1): suppose A says 'I want X' where external and publicly observable behaviour is appropriate to securing X, and where X is available. Now if A does not try to secure X, or try to do anything that he believes to be a means to securing X, we cannot conclude straight off that he does not want X and that his statement is false. In order to be able to do this, we would have to be able to say the following: (a) A has no stronger counter-wants, (b) A believes that he has the ability to succeed in securing X, and (c) A believes that he has the opportunity to secure X, or to do something that he believes to be a means to securing X. If, and only if, we were sure of these three things could we then use a behavioural criterion—that of failing to observe A

4In Chapter V, I shall defend this account, according to which a 'want'-statement entails a statement to the effect that, other things being equal, the agent will try to do or to secure something in the appropriate circumstances, against a rival account that holds that the relationship in question is one other than entailment.
trying to secure $X$—to falsify the statement. Now we can often ascertain whether or not one or more of these three things is operative in preventing $A$ from trying to secure $X$ by asking $A$ why he is not trying to secure $X$. And if $A$, in reply, does not cite anything that falls under one or more of these three heads, we would then be entitled to say that he does not want $X$, and that his statement 'I want $X$' is false. On the other hand, in order for us to verify $A$'s statement, we (or someone) should have to see $A$ trying to secure $X$, or trying to do something which is believed by $A$ to be a means to $X$. The first disjunct here is straightforward, and a simple matter of observation. The second disjunct is more complex, as we must take $A$'s beliefs into account. That is, it may be the case that $Y$ is the sole means to $X$, and we see $A$ doing $Z$. This would be a verification of the statement if we knew that $A$ believed (mistakenly, in this case) that $Z$ was a means to $X$.

Case (2): suppose $A$ says 'I want $X$' where external and publicly observable behaviour is appropriate to securing $X$, but where $X$ is not available. In this type of case, $A$ is in a privileged position to say whether his statement is true or false as long as $X$ is not available. But as soon as $X$ becomes available, then he no longer occupies a privileged position. Others are able to verify or falsify his statement in the ways indicated in case (1).

Case (3): suppose $A$ says 'I want $X$' where external and publicly observable behaviour is not appropriate to obtaining $X$. Examples of this sort of case would be when $A$ wants to think of a synonym or solve some puzzle in his head. Often a 'want'-statement of this sort can be
verified by $A$'s announcing the answer. However, such 'want'-statements are, I believe, impossible to conclusively falsify. For from the fact that $A$ does not announce the answer, we may conclude only that any one of the following is the case: (i) he tried and got the answer but wants to keep it to himself, or (ii) he tried but could not get the answer, or (iii) he did not even try. Only if we could be sure that (iii) is the case, together with the three factors indicated in case (1)—viz., he had no stronger counter-wants, he believed he had the ability to succeed, and he believed he had the opportunity—could we conclude that he did not want $X$, and that his statement that he wanted $X$ is false. Now even in cases such as this, where publicly observable behaviour is not appropriate to securing $X$, there are often visible signs of an internal struggle or trying. And by observing these signs we may on occasion verify the 'want'-statement in cases (i) and (ii). But these signs are not always present. And where they are not, though we may sometimes be tempted to say that he is not trying, we can never be sure of this. Consequently, we can never be certain that (iii) is the case. And thus, in case (3), though we can often verify a 'want'-statement, we can never conclusively falsify it.

I now want to point out an important difference between case (3), where the agent wants to think of a synonym or solve some puzzle in his head, and case (1), where the agent wants to do or to secure something, such as to climb a mountain, that requires publicly observable behaviour. Both cases have in common the feature that if the agent wants to do either of these sorts of things, then he must, other things being
equal, try to do them. But they differ in that in the case of trying to think of a synonym or solve a puzzle in one's head, the agent is not performing an action; whereas in trying to do something that requires publicly observable behaviour, such as trying to climb a mountain, the agent is performing an action. Thus trying to do or to secure something may or may not involve any action on the part of the agent. And accordingly, it would be incorrect to claim that a 'want'-statement together with a statement of certain other conditions entails an 'action'-statement. It would only be correct to claim this where what one wants to do or to secure requires him to perform an action to do or to secure it.

Now as my interest in this thesis is to give an account of the type of explanation that actions are susceptible of, I shall henceforward confine my discussion to cases where what one wants to do or to secure requires him to perform an action. Thus I shall, when I come (in Chapter V) to make use of the claim I have established earlier in this Chapter, viz., that if a 'want'-statement is truly made, then either the agent will try to do or to secure something, or there will be some countervailing wants, or the agent will not believe that he has the ability, or the agent will not believe that he has the opportunity, write it as: if a 'want'-statement is truly made, then either action will follow or one or more of the disjuncts of the 'other things being equal' clause will obtain.

I turn from the difficulties involved in verifying or falsifying a 'want'-statement, where we know the object of the agent's want, to the difficulties involved in identifying the object of the agent's want,
given that we know he wants something and observe his trying to secure something. Though the agent's trying to get something may be a publicly observable phenomenon, it may be difficult to identify what it is that he is trying to get. What I have in mind here can be illustrated as follows. Suppose we see a man pick out and purchase the largest desk in the shop. We cannot validly infer from this alone that he wanted to buy the largest desk in the shop; he may, for example, have wanted to buy a walnut desk, and, as it happened, the only walnut desk in the shop was the largest desk in the shop. And the principle of substitution known as Leibnitz's Law will not sanction a valid inference from 'He wanted to buy a walnut desk' and 'The only walnut desk in the shop is the largest desk in the shop' to 'He wanted to buy the largest desk in the shop'. The way in which he characterised the desk he was set on getting will determine which desk he wanted. And such information cannot be ascertained just by observing him pick out the desk. There is a similar difficulty in identifying the object of the agent's intention. From 'He intended to buy a walnut desk' and 'The only walnut desk in the shop is the largest desk in the shop', we cannot validly infer that he intended to buy the largest desk in the shop. As we must take how the agent conceives of his actions into account, we will never be able to use purely behaviouristic criteria to determine what it is that the agent wants or intends to do. But one's beliefs can be, and often are, known by others; this is just a fact of experience. And if the agent's action is viewed against the background of his beliefs, or character, what he wants or intends to do can always in principle, and often in fact, be determined by others.
There is a further difficulty with respect to identifying the object of the agent's want. This arises since, as I have noted, we may act directly or indirectly to secure the object of our want. Thus from the fact that we see the agent trying to secure $X$ we cannot always tell whether he wants $X$ as an end, or wants some other end for which he believes $X$ to be a means. However, we often can. There are some things that are characteristically means, such as buying plane tickets. And even in those cases in which the agent does something that can be done as an end or as a means, (e.g., buying a new car because he wants it, or buying a new car because he wants to impress people), we can often determine which is the end by taking his character, or circumstances, or beliefs, into account.

So far I have dealt with the difficulties involved in verifying or falsifying a 'want'-statement, and in identifying the object of the agent's want and intention. At this point, I wish to mention yet another difficulty: that of predicting what the agent will do given that he wants, and we know he wants, $X$. Suppose $X$ here stands for 'to get Smith to pay a debt'. Jones may, to secure this end, send Smith letters, phone him, threaten him, try to persuade him, or use other devices. And we cannot predict with certainty what he will do. This difficulty in predicting what Jones will do is also present when we say that he intends $X$. Though Jones must, in the appropriate circumstances, and other things being equal, try to get Smith to pay the debt, what means he will employ to do so is uncertain. But we may be able to get a fairly good idea of what he will do in either of these cases if we
knew enough about Jones's character, what he has done in the past in similar circumstances, his relationship with Smith, and so forth.

Let us now see if wants can be distinguished from wishes and hopes. I have already argued that we cannot be said to want to do or to secure anything without, other things being equal, trying to do or to secure something in the appropriate circumstances. If this is so, we have a way of drawing the distinction. For clearly, when we wish for something, or hope that something will be the case, we need not, other things being equal, try to do anything. But if we do not act on our wants, given that other things are equal, they are no longer our wants; they become our wishes or hopes. Thus wishes and hopes are distinguishable from wants in that whereas neither a 'wish'-statement nor a 'hope'-statement entails a statement to the effect that, other things being equal, the agent will try to do or to secure anything, a 'want'-statement does entail a statement to the effect that, other things being equal, the agent will try to do or to secure something in the appropriate circumstances.

I now want to turn to the second major issue that it is my aim to clarify in this Chapter. This concerns the descriptions under which specific desires are identifiable. As I mentioned earlier, many of the arguments designed to show that desires are not causes of actions appeal to a specific view on this issue. And as I shall be considering some of these arguments in the following Chapters, it will be important to make clear what this view is. It will also be useful to know, prior to considering these arguments, both that this view has been challenged,
and what alternative views have been offered. I shall, then, first
make clear the view concerning the descriptions under which specific
desires are identifiable that many of the arguments designed to show
that desires are not causes of actions appeal to, and then go on to
indicate the alternatives to adopting this view.

The view appealed to in many of the arguments designed to show
that desires are not causes of actions rests on the following fact
about desires. We do not simply desire; we desire to do something, or
to secure something, or that something be the case. We can also desire
to do nothing; but the sense in which we can do this, to do nothing
is the object of our desire; what we want is some free time. For every
desire we have, there is an object of that desire. Sometimes we are
not sure what it is that we want; we want something but cannot say
quite what it is. But even in such cases, there is an object of our
desire, though it is imperfectly known to us.

Some philosophers have then gone on to claim that the only way
in which we can specify our desires, and distinguish one from another,
is under some description that mentions the object, be it action, thing,
or state of affairs, that it is a desire to do, secure, or bring about.\footnote{See, for example, R. Taylor, \textit{Action and Purpose} (Englewood Cliffs, N.J., 1966), pp. 254-255, and A.I. Melden, \textit{Free Action} (London, 1961), Ch. X, esp. pp. 109-114.} That is, they claim that we are only in a position to identify a specific
desire under the descriptions 'desire to perform a certain action', 'de-
sire to secure a certain object', 'desire to bring about a certain state
of affairs', 'desire to have a certain thing', and so on. For the sake
of simplicity of exposition, their claim can be put, as I shall put it both in this and in subsequent Chapters, as that the only way in which we can identify a specific desire is under the description 'desire to perform a certain action'. But, with suitable changes, what I say about this description and the implications of it applies when other objects of our desires, (e.g., things or states of affairs), are substituted for actions.

Though this view that we are only in a position to identify a specific desire under the description 'desire to perform a certain action' is widely held, it is not universally held. Recently, it has been challenged by J.A. Fodor in his book *Psychological Explanation.* Fodor suggests two alternative ways of giving identifying references to specific desires. The first of these is that we can identify a specific desire under a description that correlates the desire with some state of affairs that is associated with it in a one-to-one fashion. This suggestion can be illustrated as follows. Suppose it happens to be the case that there is a draft in the Tower of London when and only when Smith has a desire to eat a melon. If so, we could make an unequivocal identifying reference to the desire Smith has whenever he desires to eat a melon by the use of the description 'the desire Smith has whenever the Tower of London is drafty'. And, following up this suggestion, other possible identifying descriptions of specific desires immediately come to mind. A short list of them is as follows: 'the dominant desire

---

*New York, 1968, see esp. pp. 34-35.*
I had at 10:00 this morning; 'the desire I had on seeing the melon'; 'the desire that was the effect of my not eating any lunch'. The second sort of way envisaged by Fodor of identifying a specific desire without mentioning the object of that desire, is under some neurological description.

Having thus briefly indicated some of the views concerning the descriptions under which specific desires are identifiable, I wish to defer further discussion of this topic until Chapter V. I choose to do this, because the relevance of having or not having these alternative identifying descriptions of specific desires for the question of whether or not desires are causes of actions will then be clearer than it is at this stage of the argument; and I think it best to conduct a discussion of this issue when we see what exactly turns on it.

It will now perhaps be useful to give a sketch of how the argument will proceed from here. In Chapters I and II, I have argued that to explain an agent's action by giving his reason(s) or motive(s) for action is to explain the action in terms of his desires or his desires and information. In this Chapter, I have argued that whether or not explanations in terms of the agent's desires and information (as well as, of course, explanations in terms of the agent's desires alone) are causal explanations depends on whether or not desires are causes of actions. Thus the crucial question to be answered in order to determine whether or not reason- and motive-explanations are causal explanations turned out to be 'Are desires causes of actions?' In the next Chapter, I shall begin to consider this question by reviewing some of the arguments
and considerations that have been adduced both in favour of answering it in the affirmative and in the negative. I shall then go on, in Chapter V, to give my own answer to this question. I shall argue that desires are not causes of actions, and that, hence, reason- and motive-explanations are not causal explanations.
CHAPTER IV

DESIRE AS CAUSES OF ACTIONS (I)

In this Chapter I want to raise the question 'Are desires causes of actions?'\(^1\) In dealing with this question in this Chapter, I shall first review some of the arguments and considerations that have been adduced in favour of saying that desires are causes of actions, and then turn to some of the main arguments designed to show that they are not. None of these, as we shall find, forces us to answer the question in one way or the other. I shall then go on, in Chapter V, to produce two arguments that I think will show that desires are not causes of actions.

But let me now begin the task of this Chapter by stating and examining some of the considerations that have been adduced in favour of saying that desires are causes of actions.\(^2\) The first of these is that we often use causal idioms to refer to desires when we cite them to explain actions. For example, we say things like 'My desire to get in a full day of skiing caused me to catch the early bus', or 'My desire for more speed on the downhill runs made me put on my fiberglass skis'. But perhaps the word 'because' is the causal idiom most commonly and

---

\(^1\) I remind the reader that I am using the words 'desire' and 'want' interchangeably throughout.

\(^2\) These considerations to be discussed are put forward by, among others, W.D. Gean, "Reasons and Causes", Review of Metaphysics, XIX, 4 (June, 1966), 674-676.
naturally used in these contexts, as when one says 'I took the early bus because I wanted to get in a full day's skiing', or 'I put on my fiberglass skis because I wanted more speed on the downhill runs'.

However, though we do use causal idioms in this way to refer to the agent's desires, this linguistic fact alone will not establish that desires are causes of actions. For the possibility remains that the words 'cause' and 'made' have a use other than a causal use in these contexts; they may, for example, be used in a metaphorical way to indicate the compelling (but not causal) nature of the desire. And the word 'because' cannot be taken to invariably indicate a causal relation. For we can, for example, indicate the relationship between the premises and conclusion of a valid argument, say where a conclusion r is deduced from premises p and q, by saying r, because p and q; and here we are clearly not indicating a causal relation. Thus the fact that we often use causal idioms to refer to the agent's desires is not decisive in showing that desires are causes of actions; but this fact does remain, I think, a consideration in favour of the claim.

The second consideration often adduced is that statements explaining an agent's action in terms of his desires, like causal statements, normally imply the truth of counterfactuals. For example, in saying that the bent rail caused the accident, we imply that if the rail had not been bent, then, other things being equal, the accident would not have occurred. Similarly, in saying that he did X because he wanted Y, we imply the truth of the counterfactual 'If he had not wanted Y, he would not, other things being equal, have done X'.

But this consideration is not a decisive one either; for it is not only causal statements that imply the truth of counterfactual statements. For example, the statement 'The syllogism is invalid because the middle term is undistributed' implies the truth of the counterfactual 'If the middle term had not been undistributed, then, other things being equal, the argument would not have been invalid'. But we would not, I think, say that the undistributed middle term caused the argument to be invalid. Or again, to take an example not utilising the 'because'-construction but which still yields a counterfactual, the statement 'I discharged my obligation by going to the meeting' implies the statement 'If I had not gone to the meeting, I would not, other things being equal, have discharged my obligation'. But we would not say that going to the meeting caused me to discharge my obligation; for going to the meeting is discharging my obligation, and so cannot be the cause of it. Thus this second consideration remains a prima facie one only.

The third consideration that is sometimes alleged in favour of the claim that desires are causes of actions can be stated in the form of the following argument. (1) By modifying an agent's desires we can bring about changes in his actions; (2) It is inconsistent to assert that by modifying \( x \) we can bring about changes in \( y \) and to deny that \( x \) causes \( y \); therefore, desires are causes of actions.³

³The second premise of this argument is adapted from W. Dray, Laws and Explanation in History (London, 1957), p. 94. Dray writes that we cannot consistently say 'that \( x \) does not cause \( y \) though by manipulating \( x \) we can control \( y \)'. Similar views are held by D. Gasking, "Causation and Recipes", Mind, LXIV (1955), 479–487, and H.L.A. Hart and A.M. Honoré, Causation in the Law (Oxford, 1959). The argument was put together in a slightly different form by W.D. Gean, op. cit., p. 674.
But we cannot show that desires are causes of actions by arguing in this way because premise (2) is false. A counter-example to it is as follows. Suppose I am under an obligation to go to a meeting. Now by modifying my behaviour, i.e., by going to the meeting or not, I can bring about changes in the discharge of my obligations. But we certainly would deny that going to the meeting caused me to discharge my obligation to go to that meeting. Going to the meeting is discharging my obligation to go to that meeting. These—'going to the meeting' and 'discharging my obligation to go to that meeting'—are just two descriptions of different aspects of the same bit of behaviour. The relation between going to the meeting and discharging my obligation to go to that meeting is not an empirical relation, but a conceptual one. And this is, I think, why the counter-example works against the principle stated in premise (2).

Noticing this, we may try to salvage the principle stated in (2) by restricting it to relations that are empirical. That is, we might claim that it is inconsistent to assert that by modifying $x$ we can bring about changes in $y$ and yet to deny that $x$ causes $y$, where the relation between $x$ and $y$ is an empirical relation. Perhaps, thus restricted, the principle is true; let us grant that it is. But even if true when re-formulated in this way, it cannot, as I shall now argue, be used in the argument under consideration to show that desires are causes of actions.

It is clear that we cannot rightly apply to non-empirical relations a principle whose sphere of application is limited to empirical relations. If this principle is to be used, then, the relation between desires and actions must be an empirical relation. Specifically, before
we can properly apply the principle in question and proceed to the conclusion that desires are causes of actions, it must be empirically true that if an agent's desires are modified there will be a resultant change in his actions. But the relation between desires and actions is not, as I have argued in Chapter III, an empirical relation, but a logical one, determined in virtue of the meaning of the word 'desire' (or 'want'). It is no empirical discovery that if an agent's desire to do \( X \) is changed to a desire to do \( Y \), that he will, other things being equal, do \( Y \) rather than \( X \). Rather, it is a conceptual truth: if he did not, other things being equal, do \( Y \), or if he did \( X \) rather than \( Y \), then he could not properly be said to have had the desire to do \( Y \). Thus it seems that premise (1), which states that by modifying an agent's desires we can bring about changes in his actions, expresses a conceptual, and not an empirical, relation.

If I am right in this, the argument cannot succeed. For given that the relation expressed in premise (1) is a conceptual one, the application of the principle expressed in premise (2) cannot be restricted to empirical relations without making it irrelevant to the argument. But if we leave the principle unrestricted in this way, though it would be relevant to the argument, it would also be, as I have argued, false. Thus this argument which is sometimes claimed to be a consideration in favour of saying that desires are causes of actions is not merely, as we found the first two considerations to be, indecisive; it is not a consideration at all.
Thus the two considerations we are left with to constitute a *prima facie* case for saying that desires are causes of actions are: (a) that causal idioms are often used to refer to the agent's desires when we cite them to explain actions, and (b) that, like causal statements, statements explaining an agent's action in terms of his desires normally imply the truth of counterfactuals. Now to say that these considerations are *prima facie* considerations is to say that they would be decisive if there were no other arguments or considerations against them. But there are arguments that have been put to try to show that desires are not causes of actions. These we must now consider.

The first argument I wish to consider derives from Ryle.⁴ It can be stated as follows: (1) Only events can be causes; (2) Desires are states or dispositions, not events; therefore, desires cannot be causes of actions. There are three ways in which we could refuse to accept this argument. We could agree with the view expressed in premise (1), according to which for anything to be a cause it must be an event, and deny premise (2), that asserts that desires are dispositions or states, not events. Or we could agree with premise (2), and deny that only events can be causes. Or, finally, we could deny both premises. Now if the account of wanting provided in Chapter III is correct, we cannot reject this argument in any way that involves denying premise (2); for the account I have provided there is a dispositional one. Consequently, whether or not this argument is to be rejected depends on whether or not

---

we accept the view, expressed by premise (1), that for anything to be a cause it must be an event.

This view, however, seems false. We frequently cite states and dispositions as being causes of events. For example, we say things like: the abnormally high air temperature caused the plane to crash on takeoff; the continuing presence of organisms in the blood-stream caused the fever to persist; the unusual brittleness of the aircraft's wing caused it to fall off; structural defects in the pillars caused the bridge to collapse; and so on. And in none of these cases is what is being named as the cause an event. On this matter, Urmson writes that:

... it is a mere superstition to think that only an event may be properly named as a cause. It would indeed be absurd in ordinary circumstances to give the fact that a piece of glass has the (ordinary) brittleness of glass as the cause of its breaking; but in ordinary circumstances it would be very proper to mention the (unusual) brittleness of an aircraft's wing as the cause of the wing falling off, and quite ridiculous to mention the fact that the wind was pressing against the wing in quite a normal way, if investigating the cause of an accident.5

It thus seems false to say that only events can be causes, for we have just seen that states or dispositions can also be causes. If so, Ryle's argument must be rejected. But at this point, an argument closely related to that of Ryle's can be urged. I shall now consider this argument.

The argument begins with the claim that mention of a state or disposition only gives a cause on the assumption that there was also an event that occurred not later than the event caused. For instance, in the cases of a bridge collapse or an aircraft's wing falling off, where we cite some state or disposition as being the cause, we can also indicate, or at least suggest, some preceding or simultaneous event such as the application of weight or an increase in air pressure. And though it would not be appropriate to cite the application of (normal) weight, or the (normal) increase in air pressure that an airplane encounters at various times, as being the causes of the events in question, it might be urged that mention of things like certain structural defects, or the unusual brittleness of the aircraft's wing, is only to give the causes of the events on the assumption that there were some such preceding or simultaneous events. This view seems to me to be a true one. Now one who holds this view may try to use it to show that desires are not causes of actions in the following way. He may allege that there is a relevant difference between explaining things like a bridge collapse or a wing falling off by citing some state or disposition, and explaining actions by citing some desire, (granting that desires are states or dispositions). The difference, he may urge, consists in this: whereas in the former case we are able to indicate, or at any rate suggest, some event preceding or simultaneous with the collapse or the wing falling off, we are unable to do this in the latter one. Now if it is true, as I think it is, that mention of a state or disposition only gives a cause on the assumption that there was also
some preceding or simultaneous event, and if it is further true that there is no preceding or simultaneous event in the case of desires, (construed as states or dispositions), being followed by actions, this would show that desires are not causes of actions.

But is there really a dis-analogy here? Davidson denies that there is. He accepts the view that mention of a state or disposition only gives a cause on the assumption that there was also a preceding or simultaneous event. He also accepts the view that desires are not events. But he goes on to claim that while desires are not themselves events, the coming into a state of desiring is an event. Davidson writes that:

States and dispositions are not events, but the onslaught of a state or disposition is. A desire to hurt your feelings may spring up at the moment you anger me; I may start wanting to eat a melon just when I see one; and beliefs may begin at the moment we notice, perceive, learn, or remember something.  

I think Davidson is right in claiming this. If he is, there is no relevant difference, at least on what has so far been said, between explaining a bridge collapse or a wing falling off by citing some state or disposition such as structural defects or brittleness, and explaining actions by citing some state or disposition such as a desire. In both cases there is some event that occurred not later than the event caused. In the former, the event may be something like the application of weight, or an increase in air pressure; in the latter, it may be the

onslaught of the state of desiring, or a perception, or a passing thought, that triggers the action. And if we may name a state or disposition as being the cause of a certain event in the former case, we may also do this in the latter case. Thus there seems to be nothing in the views just considered to show that desires are not, or cannot be, causes of actions.

The next two attempts to show that desires are not causes of actions that I wish to examine both appeal to some principle which, it is alleged, states a requirement that any genuine causal relation must satisfy. In each case, the argument proceeds to the conclusion that desires are not causes of actions on the ground that the relation between desires and actions fails to meet this requirement. Let us now see what these principles are.

The first principle that I shall examine is put forward, (but only to be later rejected), by A.S. Kaufman. It is as follows: (1) Two purported occurrences (or events) can be causally related if and only if their descriptions are not logically (or conceptually) related. This principle is put forward in "Practical Decision", Mind, LXXV (Jan., 1966), 41. Kaufman does not formulate the entailment of principle (1) in exactly the same way I have. According to him, the principle stated in (1) expresses a conception of the causal relationship often attributed to Hume. And, resting on this conception, is the supposition that 'if some description of an occurrence is logically related to another description of an occurrence, then those descriptions do not refer to anything which can be causally related to one another under those descriptions' (my italics). I omit this qualification because I am not sure that it makes sense to qualify 'causally related' in this way. The idea is a problematical one, and does not receive the elucidation it requires from Kaufman. And in any case, Kaufman drops it himself when he comes to illustrate the application of the principle.
principle entails that if some description of an occurrence (or event) is logically (or conceptually) related to another description of an occurrence (or event), then those descriptions do not refer to things which can be causally related to one another.

If we accept this principle, we may then try to show that desires are not causes of actions by arguing in the following way. First, we may claim that a specific desire can only be identified under the description 'desire to perform a certain action'; and that, hence, descriptions of specific desires are always logically or conceptually related to the descriptions of the actions that they are a desire to perform. We may then go on to invoke the principle stated in (1), and say that since there is always a logical or conceptual connexion between the description of a specific desire and the description of the action that it is a desire to perform, desires cannot be causes of actions.

But before we can use the principle stated in (1) in this way to preclude desires from being causes of actions, it must first be shown that no genuine causal relation can fail to satisfy the requirement stated by that principle. However, it seems possible to give counter-examples to (1). One such counter-example would be as follows: exposure to excessive sunshine is the cause of sunburn. This is a genuine and familiar causal relation; the first occurrence can be properly said to cause the second. And yet there is an obvious logical or conceptual connexion between the descriptions 'exposure to excessive sunshine' and 'sunburn'. Or again, taking a fatal dose of barbituate can be said to be the cause of death; and this in spite of the fact that there is a
logical or conceptual connexion between the descriptions 'fatal dose' and 'death'. Thus it seems that, contrary to what principle (1) states, descriptions which are logically or conceptually related can report occurrences that are causally related. Accordingly, even granting that there is a logical or conceptual connexion between the descriptions 'desiring to do \( X \)' and 'doing \( X \)', the principle stated in (1) cannot properly be used to prohibit us from saying that desires are causes of actions.

The second principle I wish to examine purports to specify a type of logical or conceptual relation that would preclude there also being a causal relation. As I shall illustrate below, several philosophers hold this principle, and appeal to it to try to show that desires are not causes of actions. The principle in question is as follows: (2) A cause must be identifiable under some description that does not mention its supposed effect. This is to say that \( A \) cannot be regarded as the cause of \( B \) unless \( A \) can be specified in some way that does not mention \( B \).

This requirement stated in (2) is different from, and narrower in scope than, the one stated in (1). It is narrower in scope and differs in the following way. According to (1), if two descriptions are logically or conceptually related, then they cannot refer to things that can be causally related. And the implication here is that the purported occurrences reported in descriptions that are logically or conceptually related cannot be causally related, whether or not alternative descriptions of those occurrences that are not logically related are available. But (2) only stipulates that unless alternative descriptions which are not logically or conceptually related are available,
then, of two descriptions that are logically or conceptually related, one cannot refer to something that can be said to be the cause of what the other refers to. Thus, according to (1), exposure to excessive sunshine cannot be said to be the cause of sunburn, nor can a fatal dose of barbituate be said to be the cause of death, for there is an internal conceptual or logical relationship between the descriptions of the alleged causes and the supposed effects. But, according to (2), exposure to excessive sunshine can be said to be the cause of sunburn, and a fatal dose of barbituate can be said to be the cause of death, for we may describe the two occurrences in both examples in a way that does not connect them logically or conceptually. Such a re-description of the first example would be as follows: exposure to ultra-violet rays over a certain frequency emanating from the sun for an excessive period of time was the cause of burns to the skin. And, of the second example, forty grains of barbituate taken at one time was the cause of death. Thus (2) states a requirement that is narrower in scope than the one stated in (1). And the counter-examples offered against (1) do not affect the truth of (2).

Let us now see how the acceptance of (2) would affect the question of whether desires can be causes of actions. If we accept (2), which holds that \( A \) cannot be regarded as the cause of \( B \) unless \( A \) can be specified in some way that does not mention \( B \), then we must admit (2a) that if a desire cannot be identified except under a description that mentions the action that it is a desire to perform, then it cannot cause the action. Now it is clear that the admission made in (2a) cannot be used to prohibit
us from regarding desires as causes of actions unless it is the case that desires cannot be identified except under descriptions that connect them with actions. But several philosophers claim that this is the case, and then go on to use the principle stated in (2) in this way, (i.e., as entailing (2a)), to deny that desires are causes of actions. For example, Richard Taylor writes that:

Apart from the ends that are their objects there is nothing to distinguish one [desire] from another. Similar remarks cannot be made about genuine causes, however, which are never characterless or indescribable apart from their effects; and from this we can conclude that desires, as they are represented in the theory before us, are not even fit candidates for causes of actions. . . .

L.W. Beck takes a similar line:

If 'desire to go to the bookstore' were causally related to 'going to the bookstore', then it would be necessary that we be able to define and identify the former without reference to the latter in order subsequently to establish a contingent relation between them. . . . The fact is that we cannot identify the 'cause' in question except by virtue of the fact that it is a desire to go to the bookstore, and thus the situation described as 'desiring to go to the bookstore' stands in a logical and not a contingent relation to going to the bookstore, _ceteris paribus_.

We also find A.I. Melden writing that:

As Humean cause or internal impression, [desiring] must be describable without reference to anything else—object desired,

---


the action of getting or the action of trying to get the thing desired; but as desire this is impossible. Any description of the desire involves a logically necessary connection with the thing desired.¹⁰

If the relation were causal, the wanting to do would be, indeed it must be, describable independently of any reference to the doing. But it is logically essential to the wanting that it is the wanting to do something of the required sort with the thing one has. Hence the relation between the wanting to do and the doing cannot be a causal one.¹¹

Now one may refuse to accept these arguments on three different grounds. Firstly, one may accept the principle stated in (2), essential to all these arguments, that holds that A cannot be regarded as the cause of B unless A can be specified in some way that does not mention B; further admit that this principle entails (2a), according to which if a desire cannot be identified except under a description that mentions the action that it is a desire to perform, then it cannot cause the action; but then go on to argue that these principles cannot be used to show that desires are not causes of actions on the ground that we can identify specific desires in ways that do not connect them with actions.

Secondly, one may refuse to accept the arguments in question by accepting the claim that specific desires are only identifiable under some description that connects them with actions, but then go on to deny that the principle stated in (2) is a principle that states a requirement that any genuine causal relation must satisfy. And if the principle

---


stated in (2) is false, no ground whatsoever has been provided for holding, as is stated by (2a), that if a desire cannot be identified except under some description that mentions the action that it is a desire to perform, it cannot cause the action.

Thirdly, and finally, one may refuse to accept these arguments by rejecting both the claim that specific desires are only identifiable under descriptions that connect them with actions, and the claim that principle (2) states a requirement that any genuine causal relation must satisfy.

Now I do not think that these arguments from Taylor, Beck, and Melden are good ones. But I do not wish to try to break them by claiming that specific desires are identifiable under descriptions that do not connect them with actions. Rather, I wish to do so by showing that the principle stated in (2), according to which \( A \) cannot be regarded as the cause of \( B \) unless \( A \) can be specified in some way that does not mention \( B \), does not specify a requirement that any genuine causal relation must satisfy.

The principle stated in (2) can be refuted if a counter-example can be found according to which one occurrence cannot be identified except under a description that connects it with its alleged effect, and yet where no conceptual incoherence is introduced in treating the occurrence as the cause of that effect. D.F. Pears offers two such
counter-examples. One of them is as follows. Fairy stories, which treat wishes as causes and describe a wish simply as concentrated willing that such and such should happen, may be incredible, but they are not conceptually incoherent. It is only a contingent fact that magic wishes do not bring about the events which are their objects. The other counter-example he offers departs from the realm of mythology; it is that fear of a particular accident may cause that accident. And even if the relevant wish and the relevant fear cannot be identified under a description that does not mention their alleged effects, no conceptual incoherence is introduced in treating these things as causes. In fact, in the case of fears, we know that such things can be causes. These counter-examples seem to me decisive against the principle stated in (2). If I am right in thinking this, the principle expressed in (2) does not state a requirement that must be met in order to be able to regard something as a cause.

Thus, even if we were to admit that specific desires are only identifiable under descriptions that connect them with their objects, and that no alternative ways of making this identification are possible, the principle stated in (2) cannot properly be appealed to to prohibit us from regarding desires as causes of actions. For if, as is the case, the violation of this principle does not prohibit us from saying that

---

the accident was caused by the fear of that particular accident, where this is the only way in which we can identify the fear, then this principle cannot rightly be used to prohibit us from saying such things as that the raising of my arm was caused by the desire to raise my arm; or, more generally, that desires are causes of actions.

In thus reviewing several of the arguments and considerations pro and con the thesis that desires are causes of actions, we have found nothing decisive either way. But on the whole, it must be admitted that the proponents of the claim that desires are causes of actions are in the stronger position. For we have found that there are two prima facie considerations in favour of saying that desires are causes of actions, and no good arguments against saying this. But though the arguments purporting to show that desires are not causes of actions considered above do not succeed in showing this, there are, I think, two arguments that will. And I shall, in the following Chapter, present these arguments.
CHAPTER V

DESIRES AS CAUSES OF ACTIONS (II)

In this Chapter, I want to present two arguments which I think will show that desires are not causes of actions. But before I turn to do this, it will perhaps be well to make clear a position of strategy I shall adopt.

For anyone concerned to show that certain things are not causes, the ideal way to proceed is by first giving an analysis of the causal relation, and then going on to argue that the thing in question fails to possess some essential feature or features revealed by that analysis. This, however, is not the way in which I shall proceed, for I neither have an analysis of the causal relation to offer, nor know of any that I care to endorse. But I do not think that this places me in a position of real weakness. For even if we have no complete analysis of the causal relation, there are, I think, some relatively uncontroversial features of the causal relation. I have in mind here, and shall exploit in what follows, two of these. The first is that the causal relation is a contingent relation. The second is that we require empirical evidence to establish the existence of any causal relation. These are the only two features of the causal relation that I need for my arguments, and I shall assume both of them. I shall appeal to the first, viz., that the causal relation is a contingent relation, in the first argument I shall present; and I shall appeal to the second, viz.,
that we require empirical evidence to establish the existence of any causal relation, in the second.

I now begin to present the first of these arguments by considering yet another principle, in addition to the two considered towards the end of Chapter IV, purporting to state a requirement that any genuine causal relation must satisfy. This principle is as follows:

(3) If $A$, under conditions $C$, is the cause of $B$, we must be able to identify $A$ in some way that is logically independent of the description 'the cause of $B$, given $C$'. I shall, in order to make convenient reference to this principle, alternatively express it in what follows as stipulating that a cause must be identifiable under some description that does not connect it causally with its supposed effect.

Now if we accept (3) as stating a requirement that must be satisfied by any genuine causal relation, then, if we are to regard desires as causes of actions, we must accept (3a) that a desire must be identifiable under some description that is logically independent of the description 'the cause of a certain action if certain other conditions are satisfied'. Or, as I shall sometimes express it, that a desire must be identifiable under a description that does not connect it causally with action.

But does (3) state a requirement that must be satisfied? Both D.F. Pears\(^1\) and J.A. Fodor\(^2\) accept principles that I take to be

---


\(^2\)*Psychological Explanation* (New York, 1968), p. 35.
equivalent to the principle stated in (3), holding that they state a requirement that any genuine causal relation must satisfy. But neither of these writers has much to say on the question of why any genuine causal relation must satisfy this principle. And I should now like to fill this gap left by their accounts by trying to show that principle (3), according to which if \( A \), under conditions \( C \), is the cause of \( B \), we must be able to identify \( A \) in some way that is logically independent of the description 'the cause of \( B \), given \( C \)', does state a requirement that any genuine causal relation must satisfy.

The causal relation is a contingent relation. Thus if it is a fact that \( A \), under conditions \( C \), is the cause of \( B \), it is a contingent fact. This means that it cannot be self-contradictory to suppose that \( A \) and \( C \) occurred and yet \( B \) did not. If, however, we cannot identify \( A \) except as falling under the description 'the cause of \( B \), given \( C \)', then it would be self-contradictory to suppose that \( A \) and \( C \) occurred and yet \( B \) did not. For in this case, in saying that \( A \) occurred, we are saying that the cause of \( B \), given \( C \), occurred; and it is self-contradictory to say that the cause of \( B \), given \( C \), occurred, \( C \) occurred, and yet \( B \) did not. And since this is self-contradictory, it could not be true that \( A \), under conditions \( C \), is the cause of \( B \). For in this case \( A \), \( C \), and \( B \) are not contingently related; and if not contingently related, could not be causally related. Thus, since the violation of principle (3), according to which if \( A \), under conditions \( C \), is the cause of \( B \), we must be able to identify \( A \) in some way that is logically independent of the description 'the cause of \( B \), given \( C \)', forces us to admit
that $A$, $C$, and $B$ do not stand in a contingent relation, given that the causal relation is a contingent relation, principle (3) states a requirement that any genuine causal relation must satisfy. And since (3a), according to which if a desire is to be a cause of action, we must be able to identify that desire under some description that is logically independent of the description 'the cause of a certain action if certain other conditions are satisfied', is entailed by (3), (3a) also states a requirement that must be satisfied if the thesis that desires are causes of actions is to be maintained.

Having thus seen that principle (3) does state a requirement that any genuine causal relation must satisfy, let us now see how Pears connects this principle with the question of whether desires can be regarded as causes of actions. The connexion here depends on how specific desires are identifiable, and what the implications of their descriptions are. Pears begins by assuming that we are only in a position to identify a specific desire under the description 'desire to perform a certain action'. Now according to Pears, the application of

---

3 Pears's views to be expounded in what follows are those he puts forward in his "Are Reasons for Actions Causes?"

4 It is, of course, a simplification to say that specific desires are only identifiable under the description 'desire to perform a certain action'. For actions are not the only objects of desires: we can also desire to have certain things, to bring about states of affairs, to secure certain items, etc. But, with suitable changes, the remarks to be made about the description 'desire to perform a certain action' and the implications of it apply when other objects of desires, (e.g., things or states of affairs), are substituted for actions. I also take this opportunity to remind the reader that, as I announced in Chapter III I would, I adopt this simplification throughout.
this description implies—and the nature and strength of this implication will be discussed shortly—the application of the description 'desire which will be followed by the execution of that action unless there is some condition obtaining, such as that the project is not believed to be feasible, or there is no occasion to do it, or the agent lacks the necessary pertinacity, etc., that will explain why the action was not performed'. This latter description Pears refers to as 'the disjunctive description'. Thus, according to Pears, the application of the description 'desire to perform a certain action' implies the application of the disjunctive description.

Now one who adopts this view just sketched, and holds that desires are causes of actions, will not merely hold that specific desires will be followed by actions, other things being equal, but that specific desires will be the cause of actions, other things being equal. Thus he will interpret the disjunctive description causally. Interpreted causally, it would read as follows: 'desire which will be the cause of that action unless there is some condition obtaining that will explain why the desire does not bring about its effect'. And I think that he must so interpret the disjunctive description; for if he did not, it would be difficult to see in what sense he maintains that desires cause actions. Thus one who accepts the view that specific desires are only identifiable under the description 'desire to perform a certain action', holds that this description implies the disjunctive description, and maintains that desires are causes of actions, will and must hold that the description 'desire to
perform a certain action' implies the description 'desire which will be the cause of that action unless there is some condition obtaining that will explain why the desire does not bring about its effect'.

It is now important to determine the nature of the implication linking the two descriptions. The implication may be one of entailment or presupposition, in Pears's sense of that latter term. By 'presupposition' Pears means that $p$ presupposes $q$ if $p$ mentions something about which we could not establish communication unless $p$ were very seldom true when $q$ was false. It will make a great deal of difference which of these two views of the implication we adopt. The acceptance of what has been said so far about the identification of specific desires only being possible under the description 'desire to perform a certain action', and about this description implying the disjunctive description, does not commit us to anything that would, according to the principle stated in (3), prohibit us from saying that desires are causes of actions. However, when we come to state—as we must—what the nature of this implication is, and the only alternatives here are entailment or presupposition, depending on which of these alternatives we adopt, we will either be allowed or prohibited by the principle stated in (3) to regard desires as causes of actions. Let me now demonstrate that this is so.

Let us suppose that the relation between the description 'desire to perform a certain action' and the disjunctive description
is entailment. Now if we hold that the description 'desire to perform a certain action' implies the disjunctive description and hold that desires are causes of actions, then we must, as I have argued, interpret the disjunctive description causally. Accordingly, if we are only in a position to identify a specific desire under the description 'desire to perform a certain action', and if this description entails the disjunctive description interpreted causally, as 'desire which will be the cause of that action unless there is some condition obtaining that will explain why the desire does not bring about its effect', then a desire will not be identifiable under a description that does not connect it causally with action. Thus we could not regard desires as causes of

---

5 It is perhaps worth pointing out that to speak, as Pears does, of one description entailing another description is to use the notion of entailment in an extended way; for entailment is a relation that is normally considered to hold between statements. And as I shall, following Pears, continue to speak of one description entailing another, a word about what is meant by saying this would be in order. We may say that one description, \( D_1 \), entails another description, \( D_2 \), if and only if \( D_1 \) can be predicated of a subject \( x \) to form a statement \( S_1 \) and \( D_2 \) can be predicated of the same subject \( x \) to form a statement \( S_2 \) such that, of the two statements so formed out of this subject and these descriptions, \( S_1 \) entails \( S_2 \). Now to say that \( S_1 \) entails \( S_2 \) is to say that it is inconsistent to assert \( S_1 \) and yet to deny \( S_2 \). Thus to say that \( D_1 \) entails \( D_2 \) is to say that it is inconsistent to predicate \( D_1 \) of a subject and yet to refuse to predicate \( D_2 \) of that subject. Let me illustrate this. The descriptions 'six feet tall' and 'less than ten feet tall' can be predicated of a common subject, e.g., John, to yield the following two statements: 'John is six feet tall' and 'John is less than ten feet tall'. And since the former statement entails the latter one, we may say that the description 'six feet tall' entails the description 'less than ten feet tall'; that it is inconsistent to predicate 'six feet tall' of a subject and yet refuse to predicate 'less than ten feet tall' of that subject. With these explanations, I hope that the extended use of the notion of entailment, as indicating a relation holding between descriptions, will create no problems.
actions on this analysis without violating the principle stated in
(3), for according to (3), a cause must be identifiable under a des-
cription that does not connect it causally with its supposed effect.
And since, as has already been argued, we cannot violate (3) without
being forced to admit that the relation in question is not a contin-
gent relation, given that the causal relation is a contingent
relation, we could not regard desires as causes of actions if the
relationship between the two descriptions is entailment.

In this case, we would be driven to this conclusion in the
following way. According to the argument now under consideration
that holds that we are only in a position to identify a specific
desire under the description 'desire to perform a certain action', and
that the relationship between this description and the disjunctive
description is entailment, if we regard desires as causes of actions,
(thus interpreting the disjunctive description causally), then a
specific desire cannot be identified in any way that does not fall
under the description 'the cause of a certain action unless there is
some condition obtaining that will explain why the desire does not
bring about its effect'. This description is equivalent to the des-
cription 'the cause of a certain action if certain other conditions
are satisfied'. Thus, accepting the argument under consideration, and
regarding desires as causes of actions, a specific desire cannot be
identified except as falling under the description 'the cause of a
certain action if certain other conditions are satisfied'. If we now
claim that the desire to perform a certain action is the cause of that
action, this claim amounts to the claim that the cause of a certain action, if certain other conditions are satisfied, is the cause of that action. And since it is self-contradictory to suppose that the cause of a certain action, given the satisfaction of certain other conditions, occurred, those other conditions were satisfied, and yet the action did not occur, the relationship between the desire, those conditions that must be satisfied if action is to follow the desire, and the action, is not a contingent relation. And if not a contingent relation, could not be a causal relation. Thus if we accept the principle stated in (3), as it seems we must, hold that a specific desire cannot be identified except under the description 'desire to perform a certain action', and hold that this description entails the disjunctive description, then it seems impossible to regard desires as causes of actions.

However, the matter is different if we hold that the description 'desire to perform a certain action' presupposes the disjunctive description. For according to this, the description 'desire to perform a certain action' is only very seldom true when the disjunctive description is false. But the possibility that the first description can be truly applied, and the disjunctive description be false, permits us to identify a specific desire under a description that does not fall under the disjunctive description. And since this is so, we may regard desires as causes of actions without the consequence of being unable to identify a specific desire except under a description that connects it causally with its supposed effect, action. And hence
the view of the causal relationship expressed in (3) gives us no
trouble. Thus if we accept (3), hold that a specific desire is only
identifiable under the description 'desire to perform a certain
action', and hold that this description presupposes the disjunctive
description, then it is possible to hold that desires are causes of
actions.

So far we have been following Pears's account, and exploring
the consequences for the question of whether or not desires are causes
of actions of holding two alternative views in conjunction with adopt­
ing the principle stated in (3), and the view that specific desires
are only identifiable under the description 'desire to perform a
certain action'. These concern the strength of the implication link­
ing the description 'desire to perform a certain action' and the
disjunctive description. But it is to be noticed that this explora­
tion only exhausts part of the dialectic of the problem. So far
nothing has been said about the consequences for the question of
whether or not desires are causes of actions of denying the view
that specific desires are only identifiable under the description
'desire to perform a certain action', and placing either of the two
alternatives indicated towards the end of Chapter III in its stead.

But before we explore the consequences of doing this latter,
let us first see how the account of the logical connexion holding
between desires and actions that I have offered in Chapter III relates
to the acceptance of the principle stated in (3) together with the
adoption of the view that specific desires are only identifiable under
the description 'desire to perform a certain action'. According to that account, if a 'want'-statement is truly made, then at least one of the following disjuncts is entailed by it: action will follow, or there will be some countervailing wants, or the agent will not believe that he has the ability, or the agent will not believe that he has the opportunity. Now since the clause 'other things being equal' was designed to cover the last three disjuncts mentioned, the claim can be put more simply as follows. Any true 'want'-statement entails a statement to the effect that, other things being equal, the agent will perform a certain action. Accordingly, we may say that the true description 'desire to perform a certain action' entails the description 'desire which will be followed by that action, other things being equal'. And if, accepting this analysis, we want to regard desires as causes of actions, we will, and I think must, interpret this latter description causally, as follows: 'desire which will be the cause of that action, other things being equal'.

At this point, I wish to remind the reader of the restriction I imposed in Chapter III that enables me to put the first disjunct here as 'action will follow', and write below that 'any true "want"-statement entails a statement to the effect that, other things being equal, the agent will perform a certain action'. I am restricting my discussion to cases where what the agent wants to do or wants to secure requires him to perform an action to do or to secure it. Without this restriction, it would be false to say these things. For, as I noted in Chapter III, one may want to do something, (e.g., conjure up a mental picture of another's face), that does not involve any action on his part.

I shall, in what follows, on occasion refer to this entailed description as 'the disjunctive description'.
Now if (a) a specific desire cannot be identified except under the description 'desire to perform a certain action', and (b) this description entails the description 'desire which will be followed by that action, other things being equal', and (c) this last description is interpreted causally, as 'desire which will be the cause of that action, other things being equal', then (d) the description 'desire to perform a certain action' entails the description 'desire which will be the cause of that action, other things being equal'. And since the relation in (d) is one of entailment, then, assuming that (a) is true, no desire can be regarded as a cause of action and be identified except as falling under the description 'the cause of a certain action, other things being equal'. If this is the case, then we cannot regard desires as causes of actions and yet identify them in such a way as to satisfy the requirement laid down in principle (3); and since principle (3) states a requirement that any genuine causal relation must satisfy, so could not regard desires as causes of actions.

We now have before us an argument designed to show that desires cannot be causes of actions. But this argument may be challenged by challenging one or both of two of its premises. Specifically, one may challenge the argument by challenging premise (b), according to which the relation between the description 'desire to perform a certain action' and the disjunctive description is entailment, or by challenging premise (a), according to which a specific desire is only identifiable under the description 'desire to perform a certain action', or by challenging both these premises. We have also seen what the
alternatives to these premises are. The alternative to adopting premise (b) is to hold that the relation in question is presupposition, in the sense that has been given that term. And there are two alternatives to adopting premise (a). One may claim that besides being able to identify a specific desire under the description 'desire to perform a certain action', we may also make the identification under either or both sorts of descriptions indicated at the end of Chapter III. The first of these was a description that correlates the desire with some state of affairs that is associated with it in a one-to-one fashion, e.g., of the form 'the dominant desire I had at 10:00 this morning'; the second was a neurological description.

As regards the dialectic of the argument, we have found that one who accepts the principle stated in (3), on which the argument is founded, also accepts premise (a), according to which a specific desire is only identifiable under the description 'desire to perform a certain action', but denies premise (b), according to which the relation between that description and the disjunctive description is entailment, claiming instead that it is presupposition, can evade the conclusion of the argument. We have not yet seen what the position is in this respect of one who accepts the principle stated in (3) together with premise (b), but who denies premise (a) in the interest of affirming that a specific desire is also identifiable under a description that correlates the desire with some state of affairs that is associated with it in a one-to-one fashion or under a neurological description.
So let us begin to consider this argument by examining premise (b), the denial of which we know will enable us to avoid the conclusion that desires are not causes of actions. And let us begin this examination by considering the claim that the relation between the description 'desire to perform a certain action' and the description 'desire which will be followed by that action, or, if not, the agent will have some countervailing wants, or fail to believe that he has the ability, or fail to believe that he has the opportunity' is presupposition and not entailment. The term 'presupposition' is used here in the special sense introduced by Pears, according to which $p$ presupposes $q$ if $p$ mentions something about which we could not establish communication unless $p$ were very seldom true when $q$ was false. Thus, when it is said that the description 'desire to perform a certain action' presupposes the disjunctive description, it is being claimed that we could not establish communication about our desires unless they could very seldom be truly ascribed to us when we do not subsequently act and when none of the conditions listed in the disjunctive description apply. Now this, I think, is true; and I think that the further implication, namely, that if there were this relation of presupposition between the two descriptions then we could establish communication about our desires, is also true. So from the point of view of what is required to establish communication about our desires, presupposition seems to be all that is necessary. Entailment would of course also be sufficient, but it would be more than is necessary.
We must now come to grips with the problem of determining which of these two is the relation obtaining between the description 'desire to perform a certain action' and the disjunctive description. Let me begin to try to do this by stating, once again, what position those who claim that the relationship is entailment are committed to hold; and, in contrast to this, to what those who claim that the relationship is presupposition are committed.

Those who hold that the relationship between the first and second descriptions is entailment must say that if the agent claims to want to do something, and, when the time for action has come, does not try to do it though he does not have any countervailing wants, or does not fail to believe that he has the ability, or does not fail to believe that he has the opportunity, then he does not really want to do it at that time. This is equivalent to saying that the first description, 'desire to perform a certain action', cannot be truly applied to the agent if, when the time for action has come, the second description, 'desire which will be followed by that action, or, if not, there will be some condition obtaining that will explain why the action was not performed', turns out to be false. So those who hold that the relationship between the two descriptions is one of entailment rest their case on the meaning of the verb 'desire' (or 'want') according to which the applicability or not of the second description is the decisive factor in determining whether or not the agent really wants to do something.

On the other hand, those who hold that the relationship between the two descriptions in question is one of presupposition, though they
must hold that the second description cannot often fail to hold when the first is applicable—for otherwise communication could not be established about our desires—cannot regard the applicability of the second description as a *deotive* check on the truth of the first description; for to do this would amount to saying that the first description entailed the second. Those who hold that the relationship is one of presupposition claim that the agent's assertion that he wants to perform a certain action is, on occasion, to be given more weight than the applicability of the second description when we are assessing whether or not an agent really wants to do something. Thus they allow that there will be times when the description 'desire to perform a certain action' can be truly applied to the agent even if, when the time for action has come, no action follows, and the agent does not have any countervailing wants, or fail to believe that he has the ability, or fail to believe that he has the opportunity.

Now it seems to me that such a case as this, where the agent states with a sincerity that cannot be doubted that he wants to do something, and yet, when the time for action has come, neither acts nor can cite any reason why he fails to perform, is logically odd. For if the agent really wants to do something, does not want to do anything more than this that would prevent him from doing that thing, believes he has the ability and opportunity to do it, and still does not try to do it, what is the force of saying that he wants to do it? In such a case, we would, I think, ask for some explanation of why the agent did not act; and in default of such an explanation, as in
this case, deny that the agent really wanted to act at the time when action was appropriate. Now assuming that the agent was really sincere when he made the statement, we could not very well deny that he had some pro-attitude towards that which he claimed to want to do. But we could, and I think would, deny that this pro-attitude was properly described as a want. In the case under discussion, this pro-attitude would, I think, be most naturally described as an idle wish.

If I am correct in these remarks about the proper application of the verb 'want' (or 'desire'), then the first description, 'desire to perform a certain action', does entail, and not presuppose, the second description, 'desire which will be followed by that action unless there is some condition obtaining that will explain why the action was not performed'. And if this is so, assuming that specific desires are only identifiable under the description 'desire to perform a certain action', we could not interpret the second description causally, and hence could not regard desires as causes of actions, without being unable to identify any specific desire under a description that does not connect it causally with its alleged effect. And since we must be able to so identify specific desires if they are to be causes of actions, we could not regard desires as causes of actions.

I now turn to the second way indicated of challenging the argument set out on page 132. This consists in challenging its first premise, namely, that a specific desire is only identifiable under the description 'desire to perform a certain action'. J.A. Fodor, in his recent book *Psychological Explanation*, has in effect challenged
This sort of argument in this way. And we may usefully begin by considering Fodor's account.

Fodor begins by specifying a requirement that he thinks any genuine causal relation must satisfy, writing as follows:

It is, of course, true that if $X$ is the cause of $Y$, then there must be some description that is true of $X$ and that is logically independent of the description 'Y's cause', and there must be some description that is true of $Y$ and that is logically independent of the description 'X's effect'.

This statement hints at something that I think is true, but it is not as clear as one could wish for. For what kind of description we require to be true of $X$ and $Y$ is not specified; and surely not just any descriptions true of $X$ and $Y$ that are logically independent (respectively) of the descriptions 'Y's cause' and 'X's effect' will enable us to say that $X$ is the cause of $Y$. For example, it might be true that $X$ was unfortunate and that $Y$ was unpleasant. And though the descriptions 'X was unfortunate' and 'Y was unpleasant' may be true of $X$ and $Y$ respectively, and are respectively logically independent of the descriptions 'Y's cause' and 'X's effect', these descriptions are surely not of the sort required to enable us to regard $X$ as the cause of $Y$. The sort of description Fodor seems to have in mind, in his statement of the requirement that he holds any genuine causal relation must satisfy, is an identifying description. That is, his position seems to be that in order for $X$ to be the cause of $Y$, we must be able to identify $X$

---

8Psychological Explanation (New York, 1968), p. 35.
under some description that is logically independent of the description 'Y's cause', and be able to identify Y under some description that is logically independent of the description 'X's effect'. And this requirement, (or, more strictly, the first part of this requirement), is, with the simplifying omission of the conditions under which X operates as a cause, identical to the one on which the argument now under consideration depends, viz., that if A, under conditions C, is the cause of B, we must be able to identify A in some way that is logically independent of the description 'the cause of B, given C'.

Now Fodor clearly thinks that this requirement can be satisfied in the case of psychological states such as motives, intentions, desires, etc., and goes on to indicate how, writing that

that demand [i.e., the requirement just stated] would be satisfied if the materially sufficient conditions for having a certain motive could be formulated in neurological terms. Indeed, the existence of any state of affairs that is associated in a one-to-one fashion with a psychological state, either by a law of nature, or by a true empirical generalization, or by a sheer accident, would permit one to make an identifying reference to that state without referring to the behaviour that it is alleged to cause.⁹

Fodor then proceeds to illustrate this suggestion. He writes (and here I adapt his remarks to the case of desires):

Suppose, for example, that it happens to be the case that there is a draft in the Tower of London when and only when Smith has a desire to eat a melon. Then the desire Smith

has when he desires to eat a melon could be unequivocally referred to without referring to the eating of the melon by employing some such form of words as 'the desire Smith has whenever the Tower of London is drafty'.

Thus Fodor claims that the view that a specific desire is only identifiable under the description 'desire to perform a certain action' is false. According to him, we can also identify a specific desire in various other ways. And we may now make a short list of descriptions under which Fodor's account suggests we may make the identification. Fodor's account suggests that, besides being able to identify a specific desire under the description 'desire to perform a certain action', we may also do so under the following sorts of descriptions:

1. The desire I have whenever the Tower of London is drafty.
2. The dominant desire I had at 10:00 this morning.
3. The desire I had on seeing the melon.
4. The desire that was the effect of my not eating any lunch.

Fodor also envisages, in addition to these descriptions that we actually have available, the possibility of our being able, at some future time, to identify a specific desire under some neurological description.

Fodor not only suggests in his account that we can identify specific desires in these ways, but also claims that the possibility of our being able to do so enables us to identify a specific desire in a way that is logically independent of the description 'the cause of Y' (where 'Y' stands for some specific action); and, consequently, that

---

there is no conceptual bar to regarding desires as causes of actions. But let us be quite clear as to what would be required in order for us to be able to identify a specific desire in a way that is logically independent of the description 'the cause of Y', and so enable us to regard desires as causes of actions. It is not enough merely to provide other identifying descriptions besides the description 'desire to perform a certain action'. In addition, these alternative identifying descriptions must be such that we can apply them without also applying the disjunctive description, i.e., the description 'desire which will be followed by the action that it is a desire to perform, other things being equal'. It would be of no help to us in regarding desires as causes of actions to say that, besides being able to identify a specific desire under the description, $D_1$, 'desire to perform a certain action', we can also identify the desire under descriptions $D_2$, $D_3$, etc., if we could not apply descriptions $D_2$, $D_3$, etc., without also applying the disjunctive description. For if we could not apply descriptions $D_2$, $D_3$, etc., without also applying the disjunctive description, these descriptions would not permit us to regard desires as causes of actions and be able to identify the desire under some description that did not connect it causally with its supposed effect, action. And, consequently, since we must be able to so identify specific desires if we are to be able to regard them as causes of actions, we still could not regard desires as causes of actions.

Perhaps an illustration will make this point clearer. Suppose I have a desire to eat a melon. We have already seen how, if we are
only able to identify this desire under the description, $D_1$, 'desire to eat a melon', and regard desires as causes of actions, we are unable to identify the desire in a way that does not connect it causally with its supposed effect. For $D_1$, 'desire to eat a melon', entails the disjunctive description, which, if we regard desires as causes of actions, we must interpret causally, as 'desire which will be the cause of my eating a melon, other things being equal'; and if we are only in a position to identify the desire under the description 'desire to eat a melon', then we are unable to regard desires as causes of actions and yet identify the desire in any way that is logically independent of the description 'the cause of my eating a melon, other things being equal'. But now the suggestion is that we can identify the desire in question in some other way, for example, under description $D_2$, 'the desire I have whenever the Tower of London is drafty'. However, if we cannot apply this description without also applying the disjunctive description, 'desire which will be followed by my eating a melon, other things being equal', the fact that $D_2$ is available will not help us to regard the desire as the cause of the action. For if we cannot apply description $D_2$ without also applying the disjunctive description, then we are still unable to interpret the disjunctive description causally, and so unable to regard desires as causes of actions, while at the same time being able to identify the desire in a way that is logically independent of the description 'the cause of my eating a melon, other things being equal'.
Thus it does not matter to the issue of whether or not desires are causes of actions if a desire can be identified under some description other than 'desire to perform a certain action'; what is important is whether or not specific desires are identifiable under descriptions which we can apply without also applying the disjunctive description. But if some identifying description satisfying this latter requirement can be found, then it would seem that the grounds so far presented would not prohibit us from regarding desires as causes of actions.

Let us now see if the alternative identifying descriptions of specific desires suggested by Fodor's account provide us with the required conceptual independence. I shall ignore for the present, but will consider later, the possibility of giving a neurological description of a specific desire. Here I only wish to consider those other descriptions suggested by Fodor's account. These were that, in addition to being able to identify a specific desire under the description 'desire to perform a certain action', we can identify a specific desire under the descriptions:

1. The desire I have whenever the Tower of London is drafty.
2. The dominant desire I had at 10:00 this morning.
3. The desire I had on seeing the melon.
4. The desire that was the effect of my not eating any lunch.

And the relevant question to be raised about these descriptions is whether or not we can apply them without also applying the disjunctive description.
There is one feature of these proposed identifying descriptions (1)-(4) that I now wish to draw attention to. When we identify a specific desire under a description of the form 'desire to perform a certain action', (e.g., 'desire to eat a melon'), one cannot go on to say that he does not know what desire is being referred to. But this is not so in the case of descriptions (1)-(4). When any of these proposed identifying descriptions is offered, it always makes sense for one to go and ask 'What desire is being referred to?' In fact, it seems to me, as I shall now try to argue, that both anyone using expressions of the sort exemplified by (1)-(4) to make an identifying reference to some specific desire he has, and anyone with whom communication is established about what desire is being referred to, must be aware of the desire under a description of the form 'desire to perform a certain action'.

The first part of this claim, viz., that the speaker must be aware of any specific desire he has under the description 'desire to perform a certain action', I take to be fairly obvious. For specific desires do not, so to speak, conceal their objects; we are immediately aware of the identity of our specific desires as being desires to perform certain actions, to secure certain things, to maintain a certain position, and so on. And this is equivalent to saying that we are

---

'Specific' here has the force of limiting the desires in question to those we can identify. Without this, or some equivalent, qualification, it would be false to claim that desires do not conceal their objects; for, as I noted in Chapter III, we can want something but yet not know what it is that we want.
immediately aware of the identity of our specific desires under a
description of the form 'desire to perform a certain action'. If so,
then when a speaker uses any of the expressions of the sort illus-
trated by (1)–(4) to make an identifying reference to some specific
desire he has, he must also be aware of the desire under another de-
scription, viz., the description 'desire to perform a certain action'.

I now turn to consider the truth of the second part of the above
claim, viz., that in order for communication to be established about
what desire is being referred to, the hearer must be aware of the
desire in question under a description of the form 'desire to perform
a certain action'. It is obvious that, in cases where a desire is
referred to by an expression that correlates it with some state of
affairs that is accidently associated with it, e.g., of the form 'the
desire I have whenever the Tower of London is drafty', no one could
possibly understand what desire is being referred to unless he
possessed some such knowledge as that whenever the Tower of London is
drafty, I have a desire to (say) eat a melon. But this is just to
require that we do identify the desire under a description of the form
'desire to perform a certain action'. Similar remarks could, I think,
be made of cases of temporal identifications of specific desires, e.g.,
of the form 'the dominant desire I had at 10:00 this morning'.

The other two examples, viz., 'the desire I had on seeing the
melon' and 'the desire that was the effect of my not eating any lunch',
seem, at first sight, to do a bit better at establishing some sort of
communication about the desire in question. But even these do not
succeed in unequivocally establishing communication about the desire in question. The description 'the desire I had on seeing the melon' does not, for if we admit that a description correlating a desire with any state of affairs that is accidently associated with it can be used to make an identifying reference to that desire, one will remain unclear as to whether or not the desire in question has anything to do with the melon, or is just correlated in a unique way with the sight of the melon. It may, for example, be the case that the desire I had on seeing the melon was a desire to go swimming. Nor would this unclarity be removed by saying, instead, 'the desire that was caused by my seeing the melon'. For there are any number of desires that may be caused in this way: I may want to eat, or paint, or squash, or feel the melon. Similarly in the case of the identifying description 'the desire that was the effect of my not eating any lunch': what the desire that is the effect of this is remains unclear. It may be, among other things, simply the desire to eat, or the desire to have an early dinner, or the desire to eat some special kind of food, e.g., a melon. And it seems that the only way in which we could remove these doubts and unclarities as to the identity of the desire in question would be to identify the desire under a description of the form 'desire to perform a certain action'. If so, then it follows that in order for communication to be established concerning the identity of the desire in question, the hearer must come to be aware of the desire under a description of the form 'desire to perform a certain action'.
For these reasons, I think that both a speaker who has some desire, and anyone with whom he establishes communication concerning the identity of the desire he has, must be aware of the desire under a description of the form 'desire to perform a certain action'. And, of course, if it is true that a speaker cannot communicate the identity of some desire of his own without identifying that desire under a description of the form 'desire to perform a certain action', he clearly could not communicate the identity of anyone else's desire unless he both knew, and identified for the hearer, that desire under a description of this form. Thus, to put it generally, we may say that whenever communication is established concerning the identity of a specific desire, both the speaker and the hearer must know that desire under a description of the form 'desire to perform a certain action'.

I shall now use this claim to try to argue that the identifying descriptions suggested by Fodor's account do not provide us with the required conceptual independence to enable us to regard desires as causes of actions. I have just argued that whenever communication is established concerning the identity of a specific desire—or, more simply, whenever a specific desire is identified—both the speaker and hearer must know that desire under a description of the form 'desire to perform a certain action'. I have also earlier argued that the application of the description 'desire to perform a certain action' entails the application of the disjunctive description, 'desire which will be followed by that action, other things being equal'. If I am correct
on these two points, then since any time a description of the sort illustrated by (1)-(4) is used to make an identifying reference to a specific desire that succeeds in identifying that desire, we also know that a particular description of the form 'desire to perform a certain action' applies, and since the application of this latter description entails the application of the disjunctive description, we cannot identify a specific desire under any description of the sort (1)-(4) and yet refuse to apply the disjunctive description. And if this is so, then, for reasons already given, descriptions (1)-(4) do not enable us to identify a specific desire under a description such that we could regard desires as causes of actions and yet be able to identify the desire under a description that does not connect it causally with its supposed effect. Thus it seems that the uniquely identifying descriptions suggested by Fodor's account do not provide us with the required conceptual independence to enable us to regard desires as causes of actions.

So far we have failed to find any description under which we can identify a specific desire that would permit us to regard desires as causes of actions. All the identifying descriptions we have considered fail to do this because we found that we could not apply any of them without also applying the disjunctive description. But there is one other possible sort of description, already alluded to, to be considered. It may be that we can identify a specific desire under a neurological description which we can apply without also applying the disjunctive description. If so, then we could avoid the difficulty we
found in regarding desires as causes of actions, and hence would not be barred from regarding desires as causes of actions on the grounds that have so far been presented.

Let us first consider what would be involved in giving a neurological description of specific desires that would be relevant to regarding desires as causes of actions. Insofar as we wish to regard desires as causes of actions, and to give a description of them solely in neurological terms, it is necessary to show that desires are identical with certain neural states. It would not be enough to show that desires can be correlated with neural states. For in this case, if we regarded the neural states as causes of actions, as we presumably would, we would have to say either that the neural states are the causes of actions, in which case desires would not be the causes, or that actions have two simultaneous causes, neural states and desires, in which case desires would not be describable solely in neurological terms. So it seems that the identity claim must be made. But to claim that desires are identical with neural states is just to affirm the truth of one version of the Identity Thesis. I say 'one version', for the label 'the Identity Thesis' does not uniquely identify a specific thesis. So far as I have been able to determine, there are two different theses that go under this name. These differ not in the ways in which they try to establish a particular conclusion, but in the conclusions they try to establish. And only one of these, held in a particular way,—the reason for this qualification to emerge shortly—seeks to show that desires are identical with neural states.
I shall now try to locate the particular version of the Identity Thesis that will have to be maintained if we are to have the relevant sort of neurological descriptions of specific desires. To do this, I shall begin by briefly stating the two theories that both go under the name of 'the Identity Thesis'. The first theory to be stated may be called the Reductive Identity Thesis. A proponent of this theory does not wish to deny the existence of mental phenomena; he only wishes to deny that they constitute an irreducibly different sort of phenomena from physical phenomena. According to the Reductive Identity Thesis, mental phenomena are contingently and strictly identical with, and reducible to, certain neurological states of the organism.

According to the second theory, the relation between mental phenomena and physical phenomena is 'not strict identity, but rather the sort of relation which obtains between, to put it crudely, existent entities and non-existent entities when reference to the latter once served (some of) the purposes presently served by reference to the former—the sort of relation that holds, e.g., between "quantity of caloric fluid" and "mean kinetic energy of molecules". Just as we are now prepared to identify 'what used to be called quantity of caloric fluid' with molecular motion of a certain sort, so this sort of Identity Theorist maintains that we may also be able to identify 'what people now call mental phenomena' with certain sorts of neurological states. And, further, once we make this identification, just as we are prepared

---

in the former case to say that there is no such thing as caloric fluid, so we may also go on to say in the latter case that there are no such things as mental phenomena. This theory may thus be called the Eliminative Identity Thesis.\(^\text{13}\)

Now of these two versions of the Identity Thesis, only one who holds the reductive version can claim that desires are causes of actions. And hence only a Reductive Identity Theorist is in a position to try to provide neurological descriptions of specific desires that would be relevant to regarding desires as causes of actions. A proponent of the Eliminative Identity Thesis cannot claim that desires are causes of actions, for, on that theory, the existence of mental phenomena such as desires and sensations is denied. And not being able to claim that desires are causes of actions, an Eliminative Identity Theorist is not in a position to try to provide neurological descriptions of specific desires that would be relevant to regarding desires as causes of actions.

\(^{13}\)In presenting these two versions of the Identity Thesis, I have talked undiscriminatingly about the reduction or elimination of mental phenomena. But it should be noted that this is an oversimplification: not all proponents of these theories hold that their accounts apply to the whole sphere of mental phenomena. For example, U.T. Place ("Is Consciousness a Brain Process?", *British Journal of Psychology*, XLVII (1956), 44-50) and J.J.C. Smart ("Sensations and Brain Processes", *Philosophical Review*, LXVIII (1959), 141-156), who are proponents of the Reductive Identity Thesis, do not hold this theory with respect to volitional concepts (e.g., intending, desiring) or cognitive concepts (e.g., knowing, believing), but only with respect to other mental phenomena concepts such as sensation, consciousness, experience, and mental imagery. And R. Rorty (Op. cit.), who is perhaps the most explicit proponent of the Eliminative Identity Thesis, only argues for an elimination of sensations. I have, however, extended both theories to cover the whole field of mental phenomena solely in the interest of bringing the only relevant sorts of concepts to the discussion at hand—volitional concepts—inside the compass of those theories.
Thus we may exclude, as being irrelevant to the question at hand, the eliminative version of the Identity Thesis.

Accordingly, we may say that if we are to have neurological descriptions of specific desires that would be relevant to regarding desires as causes of actions, the Reductive Identity Thesis must be maintained. But this is not yet precise enough. For the Reductive Identity Thesis can be held, as it sometimes has been, (e.g., by Smart and Place), in such a way as to be inapplicable to volitional concepts such as wanting and intending. And, held in this way, the truth of the Reductive Identity Thesis would clearly be of no help to one who wishes to give neurological descriptions of specific desires. If we are to have neurological descriptions of specific desires that would be relevant to regarding desires as causes of actions, we should have to support the Reductive Identity Thesis when that theory is interpreted as seeking to identify specific desires with neural states. But if it is possible to support the Reductive Identity Thesis in this form, then it would also seem possible to identify specific desires under purely neurological descriptions which we could apply without also applying the disjunctive description. And if such identifying descriptions of specific desires are possible, nothing that has been said so far would prohibit us from regarding desires as causes of actions; for then specific desires would be identifiable under descriptions such that we could treat desires as causes of actions, and yet be able to identify the desires in ways that do not connect them causally with their supposed effects.
It thus becomes a matter of some importance to determine whether or not the Reductive Identity Thesis can be maintained in such a way as to support the contention that specific desires are identical with neural states. So let us now see whether or not this can be done. I shall argue that it cannot be.

One can try to show that a philosophical position such as that of the Reductive Identity Theorist is mistaken either by examining and criticising the arguments such a theorist uses to try to establish his conclusion, or by moving directly to a consideration of the conclusion to be established, and trying to show that it cannot be established in any way. The two arguments against the Reductive Identity Theory that I shall discuss in what follows both try in the latter way to show the theory to be mistaken.

The first argument I shall state and examine is one that Fodor describes as 'perhaps the most important argument for the view that no statement of the form "x is y" could be significant where x is a mental term, y is a physiological term, "is" means identity, and all terms bear their current senses'.\textsuperscript{14} This argument begins with a statement of a condition that it is alleged any genuine identity statement must satisfy. This condition is embodied in what is known as the Law of Transferable Epithets, according to which if x is identical with y, then any predicates meaningfully applicable to x must also be meaningfully applicable to y. Or, to put it another way, if x is identical with y,

\textsuperscript{14} Op. cit., p. 100.
and if \( Fx \) makes sense (is linguistically possible), then \( Fy \) must also make sense (be linguistically possible).  

The proponent of this argument then alleges that mental phenomena such as desires and after-images can be described in ways in which it makes no sense to describe physical phenomena such as neural states. For example, he urges that while it makes sense to describe desires as 'intense', 'compelling', 'weak', 'fluctuating', etc., and after-images as 'circular', 'green', 'dim', 'fading', etc., it does not make sense to describe neural states in these ways. And conversely, he alleges that physical phenomena such as neural states can be described in ways that would be inappropriate to the characterisation of mental phenomena such as desires and after-images. For example, he claims that while it makes sense to describe neural states as occupying a certain spatial location, no sense attaches to saying that desires and after-images have a certain spatial location. Yet, the argument runs, if mental phenomena were identical with a certain sort of physical phenomena such as neural states, it must make sense to mutually transfer these sorts of predicates; but since it makes no sense to do this, mental phenomena cannot be identical with a certain sort of physical phenomena.

In effect, the proponent of this argument charges the Reductive Identity Theorist with holding a theory that involves the commission of what Ryle has termed a 'category mistake'. That is, he alleges that since the requirement stated by the Law of Transferable Epithets must

\[ \text{This formulation of the Law of Transferable Epithets is Fodor's, } \textit{ibid.} \]
be satisfied by any genuine identity statement, if we are to identify mental phenomena with a certain sort of physical phenomena, we must be able to transfer predicates applicable to mental phenomena to expressions concerning that sort of physical phenomena, and predicates applicable to that sort of physical phenomena to expressions concerning mental phenomena. But, he continues, when we try to transfer these predicates, we find that we are applying predicates that belong to expressions of one logical category to expressions that belong to a different logical category, and so end up talking nonsense. Thus, it is alleged, the Reductive Identity Thesis cannot be true.

This argument, I think, presents a serious difficulty for the Reductive Identity Theorist. But there are some replies open to the Reductive Identity Theorist, which I shall now examine. The argument depends on the truth of two premises: (1) The Law of Transferable Epithets states a requirement that any genuine identity statement must satisfy, and (2) Statements such as 'Neural state \( N \) was intense' and 'The desire occurred in such-and-such a lobe of the cerebral cortex' are nonsense statements. Both these premises can be challenged.

The Reductive Identity Theorist may try to meet this argument by challenging premise (2), and claiming that though expressions like 'The desire occurred in such-and-such a lobe of the cerebral cortex' or 'Neural state \( N \) was intense' are odd, they are not nonsensical. One who takes this line tries to resist the assimilation of these expressions to the class of statements that the proponent of the Argument from Transferable Epithets claims they belong, and tries instead to
assimilate them to another class. That is, a proponent of the Argument from Transferable Epithets tries to assimilate these expressions to the class of clear nonsense statements such as 'Saturday is tired' or 'The armchair is wise'. But one who claims that such statements are only odd, not nonsense, tries to assimilate them to the class of statements which are a bit odd-sounding, but clearly not nonsensical, such as 'NaCl is a tasty relish' or 'H₂O is a refreshing drink'. Thus the proponent of the Argument from Transferable Epithets and the Reductive Identity Theorist who wishes to resist that argument in this way are in agreement on the point that the expressions in question are odd. But whereas the former goes on to claim that this oddity constitutes nonsense, the latter denies this.

How can we decide on which side of the sense/nonsense line these statements fall? I find this question difficult to answer, owing to the lack of any precisely drawn dividing line between the two classes of statements. In the case of some expressions such as 'Saturday is tired', we can clearly say that they are not only odd, but nonsense; and in the case of others, such as 'NaCl is a tasty relish', we can clearly say that they are only odd, not nonsense. But the sorts of expressions now under consideration such as 'The desire occurred in such-and-such a lobe of the cerebral cortex' and 'Neural state N was intense' seem to lie between these two flanking clear cases. And it is not clear to which class of expressions these belong. What is needed is a logic of nonsense: we need some criterion or criteria by which we can judge whether or not an oddity constitutes nonsense. But lacking this, as we
do, it is hard to see what conclusion ought to be drawn from the fact
that the statements in dispute are odd-sounding. The proponent of
the Argument from Transferable Epithets may urge that one ought not
to hold theories that commit one to saying things that may be, as
they seem to many to be, nonsense; but the Reductive Identity Theorist
who seeks to resist the Argument from Transferable Epithets in the way
just sketched, may equally forcefully reply that until the statements
alleged to be nonsense are shown to be in fact so, he need not give
up the theory that commits him to them.

This possible reply of the Reductive Identity Theorist to the
Argument from Transferable Epithets I think effectively dulls the edge
of that argument. But it is not as conclusive a reply as one might
wish for. For while it casts doubt on one of the premises of that
argument, it does not establish that that premise is actually false.
However, there is a stronger line a Reductive Identity Theorist can
take to resist the Argument from Transferable Epithets which can be
pressed with more success. I now turn to consider this line.

A Reductive Identity Theorist may try to meet the Argument from
Transferable Epithets by challenging premise (1), and denying that the
Law of Transferable Epithets states a requirement that must be satis-
"fied by any genuine identity statement. According to this Law, if \( x \)
is identical with \( y \), and if \( Fx \) makes sense, \( Fy \) must also make sense.
But there are instances of identity statements that do not seem to be
impugned by failing to meet this requirement. For example, we say that
the temperature of a gas is identical with the mean kinetic energy of
of the gas molecules. However, though we can sensibly speak of the temperature of the gas being 80°C, it does not seem to make clear sense to say that the mean kinetic energy of the gas molecules is 80°C. And even granting that it is nonsense to say the latter, we would not, I think, conclude from this that the identity statement in question is somehow unjustified. The fact is that we sometimes allow as legitimate what may be termed, (to use Cornman's phrase), 'cross-category' identities of this sort. In the light of this, the Reductive Identity Theorist may contend that the Argument from Transferable Epithets poses no problem for his theory. For if the Law of Transferable Epithets does not state a requirement that any genuine identity statement must satisfy, that law cannot be rightly used in an argument to show that an alleged identity statement cannot in fact be one.

This shows, I think, that the Argument from Transferable Epithets will not defeat the Reductive Identity Theorist. But even though he can meet that argument in this way, he is not completely free from difficulties. For he must now try to show that the identity he wishes to claim is an instance of a legitimate cross-category identity. I shall now present an argument designed to show that he cannot do this in the case of volitional concepts. I shall present this argument with

---


17 These two counters to the Argument from Transferable Epithets just considered are not the only ones, but they are I think the most plausible. Fodor produces a third counter (*op. cit.*, pp. 105-106) in addition to slightly different versions of the two I have presented.
reference to desires; but a similar argument could be generated to cover intentions.  

Let me begin by re-stating the position of the Reductive Identity Theorist. The Reductive Identity Theorist does not wish to deny any of the familiar facts about mental phenomena; he only wishes to deny that they constitute an irreducibly different sort of phenomena from physical phenomena. It is his view that we can account for all so-called mental phenomena in terms of certain physical states of the organism; and, specifically, that the so-called mental phenomena are contingently and strictly identical with, and reducible to, certain physical phenomena, which are usually supposed to be sorts of neurological phenomena. Thus, to limit the discussion to the relevant case at hand, the case of desires, the Reductive Identity Theorist (presupposing, of course, that such a theorist holds the Reductive Identity Theory in such a way as to be applicable to volitional concepts) will claim that desires are contingently identical with, and reducible to, certain neural states.

Now when one type of phenomena, A, is reduced to, by way of being identified with, another type of phenomena, B, two conditions must be satisfied. (1) We must be able to explain all phenomena explained in terms of A (the phenomena to be reduced) in terms of B (the phenomena to which A is reduced), and (2) the explanations in terms of B must be more

---

18 The argument that follows is adapted from one used in a different context by N. Malcolm, "The Conceivability of Mechanism", *Philosophical Review*, LXXVII (1968), 45-72. See esp. secs. 1-7.
basic than the explanations in terms of \( A \). In order for this second condition to be satisfied, the laws invoked in \( B \)-type explanations must be more basic than the laws invoked in \( A \)-type explanations. And to say that \( B \)-laws are more basic than \( A \)-laws is to say that the regularities reported in \( A \)-laws are dependent on the regularities reported in \( B \)-laws, but that the converse does not hold. If we could not satisfy condition (1), i.e., give explanations of phenomena explained in terms of \( A \) in terms of \( B \), we could not claim that \( A \) is identical with \( B \). And if we could not satisfy condition (2), i.e., show that explanations in terms of \( B \) are more basic than explanations in terms of \( A \), we could not claim that \( A \) is reducible to \( B \).

If this is right, then when one claims that desires are identical with, and reducible to, certain neural states, he must be able to (a) explain phenomena explained in terms of desires in terms of neural states, and (b) show that the explanations in terms of neural states are more basic than the explanations in terms of desires. Now with respect to (a), we commonly explain actions in terms of the agent's desires. Consequently, if desires are to be identified with certain neural states, we must be able to explain actions in terms of those neural states. And these explanations, according to (b), must be shown to be more basic than the explanations in terms of the agent's desires.

The argument against the Reductive Identity Theory I shall now present grants that the first requirement can be satisfied, but questions that the second can be. In order to see whether or not this second requirement can be satisfied, we first need to determine what sort of law
explanations of actions in terms of the agent's desires appeal to, and what sort of law explanations of actions in terms of the agent's neural states appeal to. With an eye to doing this, let me now set out the forms of the two competing explanations.

When we explain an agent's action by citing some desire of the agent's, as when we say that $A$ tried to do $X$ because he wanted to, the form of this explanation can be set out as follows:

I. 1. Whenever an agent wants to do $X$, then he will, other things being equal, try to do $X$.

2. $A$ wanted to do $X$, and other things were equal.

Therefore, $A$ tried to do $X$.

In opposition to this, if we were to explain an agent's action by citing some neural state, saying that $A$ tried to do $X$ because he was in neural state $N$, the form of this explanation could be set out as:

II. 1. Whenever an agent of structure $S$ is in neural state $N$, then he will, other things being equal, try to do $X$.

2. $A$ (an agent of structure $S$) was in neural state $N$, and other things were equal.

Therefore, $A$ tried to do $X$.

A Reductive Identity Theorist must now try to show that explanations of the sort illustrated by II are more basic than those of the sort illustrated by I.\textsuperscript{19} And to do this, he must show that the law

\textsuperscript{19}It will be noticed that I have chosen simplified cases of explanations, viz., explanations just in terms of some desire of the agent's, to illustrate the two competing forms of explanations. More
appealed to in II is more basic than the law appealed to in I. This involves showing that the regularity stated by the law appealed to in I, viz., whenever an agent wants to do \( X \), then he will, other things being equal, try to do \( X \), is dependent on the regularity stated in the law appealed to in II, viz., whenever an agent of structure \( S \) is in neural state \( N \), then he will, other things being equal, try to do \( X \), but that the converse does not hold.

It is, however, impossible to show this. For the law appealed to in I is, as I argued in Chapter III, an \( a \) \( priori \) law, made out solely in virtue of the meaning of the verb 'want' (or 'desire'). But the law appealed to in II is a contingent law. And regularities embodied in \( a \) \( priori \) laws, i.e., regularities determined solely by consideration of the meanings of certain terms, cannot be dependent on regularities embodied in contingent laws, i.e., regularities determined by empirical investigation. Thus the neurological law appealed to in II could not be more basic than the \( a \) \( priori \) law appealed to in I. It follows from this that the sort of explanation illustrated by II could not be more basic than the sort of explanation illustrated by I. And if not, it would seem that a Reductive Identity Theorist cannot make out the claim that desires are reducible to, and identifiable with, certain neural states. For in this case one of the necessary conditions, any reductive frequently, reason-explanations are not just in terms of some desire the agent has, but in terms of his desires and information. However, the point I shall now make concerning these two forms of explanations also applies to explanations in terms of the agent's desires and information.
account must satisfy cannot be satisfied, namely, that explanations in
terms of the phenomena to which some other phenomena are reduced must
be more basic than explanations in terms of the phenomena to be reduced.

It would perhaps be in order here to state a limitation of the argument just produced. If sound, it will only show that volitional concepts such as wanting and intending cannot be identified with, by way of being reduced to, certain neural states. It will not show that other mental concepts such as consciousness, experience, sensation, and mental imagery are irreducible to, and unidentifiable with, neural states; for there are no a priori laws pertaining to these sorts of phenomena. So for all that has been shown, one could be a Reductive Identity Theorist about some mental concepts, viz., the last sort mentioned above. But in any case, if my argument tending to the conclusion that desires cannot be identified with, and reduced to, certain neural states is sound, we cannot give neurological descriptions of specific desires that would be relevant to regarding desires as causes of actions.

Thus it seems that we cannot avoid the conclusion of the argument set out on page 132 by denying its first premise, namely, that a specific desire is only identifiable under the description 'desire to perform a certain action'. For the only alternatives to holding this position seem to be to claim that we may also identify a specific desire under a description that correlates the desire with some state of affairs that is associated with it in a one-to-one fashion, (e.g., of the form 'the dominant desire I had at 10:00 this morning'), or that we may do so
under a neurological description. And, on examining these alternatives, we found, in the case of the former, that such identifying descriptions will not help; and, in the case of the latter, that such descriptions are not, in principle, available.

And this finding, viz., that we cannot avoid the conclusion of the argument by denying its first premise, taken in conjunction with the earlier finding that the relation between the description 'desire to perform a certain action' and the disjunctive description is, as the second premise has it, entailment and not presupposition, yields the consequence that we cannot escape the conclusion that desires cannot be causes of actions.

I have just completed one argument designed to show that a negative answer must be given to the question 'Are desires causes of actions?' But even if my argument is mistaken at some point, or can be overcome in some way, there is, I think, another argument that will establish the thesis that desires are not causes of actions. I shall now present this argument.

The basis of the argument now to be presented consists in the fact that the grounds we require to claim that $A$, under conditions $C$, will cause $B$ are totally different from the grounds on which we claim that a desire to perform a certain action, other things being equal, will be followed by that action. In order to say that $A$, under conditions $C$, will cause $B$, we require some empirical evidence. This evidence may consist in the experience of events of type $A$ being regularly followed by events of type $B$ under conditions $C$. Or perhaps it may
consist in the experience of a single instance of A and C being followed by B under carefully controlled conditions. Or it may even be of an indirect sort; for example, we might predict that compound X will cause sleep on the basis of knowledge of similar compounds. That is, we might know that compound X is closely related to nitrous oxide, and that nitrous oxide, under certain conditions, causes sleep; and on the basis of this predict that compound X, under certain conditions, will cause sleep. I do not contend that the evidence of the above three sorts is sufficient for claiming that A, under conditions C, will cause B. Perhaps an observed regularity of sequence is not enough to entitle us to say that A, under conditions C, will cause B; perhaps a single observation, however careful, would not be sufficient; perhaps we could not argue from our knowledge of similarities. All I need for my argument is that we cannot, totally independent of experience, say that A, under conditions C, will cause B; that some empirical evidence (the nature and extent of which need not be specified) is required for the claim. And this is all that I am claiming here.

If I am correct in this, then before we can make any causal statement to the effect that A, under conditions C, will cause B, we require empirical evidence. Thus if we can say that a particular desire, other things being equal, will be followed by a particular action without any empirical evidence whatsoever, then this statement will not be a causal one. And this just so happens to be the case. No empirical evidence is needed for, nor appropriate to, the claim that if we want to do something we will, other things being equal, try to do it in the
appropriate circumstances. This claim, as I have argued in Chapter III, is true in virtue of the meaning of the verb 'desire' (or 'want'); the connexion between desires and actions is a logical or conceptual connexion, made out totally a priori. Accordingly, since empirical evidence is required to say that \( A \), under conditions \( C \), will cause \( B \), and since no empirical evidence is required to say that a desire to do \( X \), other things being equal, will be followed by the doing of \( X \), desires are not causes of actions.

But, at this point, one might make the following objection. Granted that in the case of statements expressing ordinary causal relations empirical evidence is required, and that in the case of statements expressing the relation between desires and actions it is not, this fact only shows that the relation between desires and actions is not a causal relation of the ordinary sort. It does not show that the relation is not a causal one in any sense; for the possibility remains that it is an extra-ordinary sort of causal relation.

But if one were to maintain this position, holding that empirical evidence is required before we can make any causal statement except when we come to relations such as that between desires and actions, I think the difference between his position and mine turns out only to be one of terminology. That is, one who holds that there is a sense in which the relation between desires and actions is a causal relation, and I who hold that the relation between desires and actions is not a causal one, agree that the relation in question is not a causal relation in the ordinary sense. We only disagree on whether or not the relation between
desires and actions ought to be called a causal relation in some sense. My reason for saying that it ought not, is just that it is misleading to do so. There is, after all, an important difference between ordinary causal relations and the relation between desires and actions; and this difference ought not to be obscured in the way that calling them both causal relations (though in different senses) would tend to do. Nothing is gained by calling them both causal relations, but there is, I think, a loss in clarity. Thus I suggest that we ought to mark the difference between the two, and do so by denying that the relation between desires and actions is a causal relation.
CHAPTER VI

CONCLUDING REMARKS

The findings of Chapters I–V may now be drawn together to yield an answer to the question 'What kind of explanation are we giving when we explain an agent's action in terms of his reason(s) or motive(s) for action?' We found in Chapter I that when we explain an agent's action by giving his reason or reasons for action we are explaining the action in terms of his desires or his desires and information. We found in Chapter II that when we explain an agent's action by giving his motive or motives for action we are explaining the action in terms of his desires and information. Now clearly, when we explain an agent's action just in terms of his desires, whether or not this is a causal explanation depends on whether or not desires are causes of actions. And I argued in Chapter III that whether or not explanations in terms of the agent's desires and information are causal explanations also depends on whether or not desires are causes of actions. Thus the crucial question to be answered to determine whether or not reason- and motive-explanations are causal explanations turned out to be 'Are desires causes of actions?' In Chapters IV and V, I took up this question, and we found that desires are not causes of actions. And with this finding, we must conclude that reason- and motive-explanations are not causal explanations, but a completely and irreducibly different sort of explanations altogether.
But one might now feel that this conclusion is in some way unsatisfying. Davidson, for example, writes that

One way we can explain an event is by placing it in the context of its cause; cause and effect form the sort of pattern that explains the effect, in a sense of 'explain' that we understand as well as any. If reason and action illustrate a different pattern of explanation, that pattern must be identified.¹

And later, that

If, as Melden claims, causal explanations are 'wholly irrelevant to the understanding we seek' of human actions then we are without an analysis of the 'because' in 'He did it because . . . ', where we go on to name a reason.²

Now in saying this, Davidson could not mean, or could not sensibly mean, that if we deny that reasons are causes, then we do not or cannot explain a particular action by citing some desire or some desire and some related item of information that the agent has. For if I say I am stopping the car because I am hungry and know that there is a restaurant nearby I am explaining my action; and what I say does not cease to be an explanation if I go on to deny that the 'because' indicates a causal connexion. In one sense, the issue of whether or not reason-explanations are causal explanations does not affect our understanding of actions in terms of the agent's reason for action.


²Ibid., p. 230.
Specifically, however we regard the status of reason-explanations, i.e., as being causal or non-causal, this will not, on an everyday level, add to or detract from the intelligibility of explanations we give of human actions in terms of the agent's reason for action. For these explanations are perfectly intelligible whether or not we go on to assert or to deny that they are causal explanations.

But on another, and deeper, level, it may be claimed, this issue is important. For one may argue, and I take it that this is the force of Davidson's comments, that if reason-explanations are not causal explanations, the connexion between the agent's reason and his action is in some way mysterious; but if we hold that such explanations are causal explanations, thus holding that the connexion is a causal one, this connexion becomes understandable and so no mystery arises. But this is, I think, an illusion. Why should the labelling of a connexion as 'causal' be supposed to add to our understanding of that connexion? We have, it is true, supplied a name for the connexion that we often apply to connexions between events. But this labelling of the connexion does not seem to me to make any advance, as regards our understanding of the connexion, over simply saying that the agent had a reason for doing $X$ and so did $X$ where one denies that the connexion is a causal one. We could, if we wished, to dispel the air of negativism that attends the non-causal theorist's denial, supply a name for this connexion; we might call it, say, a 'rational' connexion. This, it is true, contributes nothing to our understanding of the connexion; but neither, I submit,
does calling the connexion 'causal' do so. The history of analyses of causation, both remote and recent, testifies to the fact that there is no wide agreement over what is meant by a causal connexion. This being so, there seems to be no gain in assimilating reason-explanations to causal explanations. And it seems to me that any way in which what I have termed a 'rational' connexion may be called 'mysterious', so may a causal connexion be called.  

These remarks are intended to show that there is nothing special about causal explanations that justifies our giving them a pre-eminent position as regards their familiarity or explanatory power. If they do this, then at least one, and I think the major, possible source of dissatisfaction with my conclusion that reason- and motive-explanations are a completely and irreducibly different sort of explanations from causal explanations will have been removed. And the removal of this source of dissatisfaction should make for an easier acceptance of the conclusion I have argued for.

---

3Similar remarks to those I have made in this and the preceding paragraph are offered by R.J. Richman, "Reasons and Causes: Some Puzzles", Australasian Journal of Philosophy, XLVII (1969), 43.
## Bibliography

Bibliography entries are collected under the following heads:

<table>
<thead>
<tr>
<th></th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>WANTS AND DESIRES</td>
</tr>
<tr>
<td>II</td>
<td>INTENTION</td>
</tr>
<tr>
<td>III</td>
<td>ACTION</td>
</tr>
<tr>
<td>IV</td>
<td>REASONS AND CAUSES</td>
</tr>
<tr>
<td></td>
<td>(a) Historical</td>
</tr>
<tr>
<td></td>
<td>(b) Favourable to the Distinction</td>
</tr>
<tr>
<td></td>
<td>(c) Critical of the Distinction</td>
</tr>
<tr>
<td></td>
<td>(d) Other Relevant Discussions</td>
</tr>
<tr>
<td>V</td>
<td>MOTIVES</td>
</tr>
<tr>
<td>VI</td>
<td>PRACTICAL REASONING AND REASONS FOR ACTION</td>
</tr>
<tr>
<td>VII</td>
<td>ANTHOLOGIES</td>
</tr>
</tbody>
</table>
I. WANTS AND DESIRES

Abelson, R. "Because I Want To." Mind, LXXIV (1965), 540-553.


———. "Wants, Actions, and Causal Explanation." Intentionality, Minds, and Perception, ed. H.N. Castaneda, Detroit, 1967, pp. 301-341. See also the comments on this paper by Keith Lehrer and a rejoinder by Alston in the same volume.


Pashman, J. "Raziel Abelson on 'Because I Want to'." *Mind*, LXXVII (1968), 581.


Russell, B. *Analysis of Mind*. London, 1921. Ch. III.


II. INTENTION


Gibson, Q. The Logic of Social Enquiry, New York, 1960. Ch. IV.


III. ACTION


———. "Mention and Use as Applied to Nonlinguistic Actions." Philosophical Studies, XIX (1968), 5-12.


Alston, W.P. "Wants, Actions, and Causal Explanation." Intentionality, Minds, and Perception, ed. H.N. Castaneda, Detroit, 1967, pp. 301-341. See also the comments on this paper by Keith Lehrer and a rejoinder by Alston in the same volume.


Dowling, R.E. "'Can an Action Have Many Descriptions?" *Inquiry*, X (1967), 447-448. See the reply by A.B. Cody in the same issue.


Hobbes, T. Leviathan. 1651. Esp. Pt. I, Ch. VI.

Holmes, O.W. The Common Law. London, 1911. Ch. II.


Locke, J. *An Essay Concerning Human Understanding.* 1690. Bk. II, Ch. XXI.


Mill, J.S. *A System of Logic*. 1843. Bk. III, Ch. V, Sec. II.


Oakeshott, M. "Rational Conduct." *Cambridge Journal*, IV.


Shaffer, J.A. *Philosophy of Mind.* Englewood Cliffs, N.J., 1968, Ch. V.


White, A.R. The Philosophy of Mind. New York, 1967, Ch. VI.

Wilkins, B.T. "The Thing To Do?" Mind, LXXIV (1965), 89-91. Some comments on Dray's Laws and Explanation in History.


IV. REASONS AND CAUSES

a) Historical:


Mill, J.S. *A System of Logic*. 1748. Bk. VI, Ch. II.

b) Favourable to the distinction:


Taylor, R. *Action and Purpose.* Englewood Cliffs, New Jersey, 1966, Ch. XVI.


c) Critical of the distinction:


Goldberg, B. "Can a Desire Be a Cause?" Analysis, XXV (1965), 70-72.


Hancock, R. "Interpersonal and Physical Causation." Philosophical Review, LXXI (1962), 369-376.


d) Other relevant discussions:

Braithwaite, R.B. *Scientific Explanation.* Cambridge, 1955. Ch. X.


Ginet, C. "Can the Will Be Caused?" *Philosophical Review,* LXXI (1962), 49-52.


Thalberg, L. "Do We Cause Our Own Actions?" *Analysis,* XXVII (1967), 196-201.

V. MOTIVES


Aristotle. *De Anima*. Bk. III, Chs. IX-XI.

———. *Ethica Nicomachea*. Bk. III, Chs. III-IV.


Nowell-Smith, P.H. *Ethics.* Baltimore, 1954. Ch. IX.


———. "Motives and Motivation." *Philosophy, XXXI* (1956), 117-130. The main arguments of this appear in Ch. II of his *The Concept of Motivation*.


VI. PRACTICAL REASONING AND REASONS FOR ACTION


——. *De Motu Animalium.* Ch. VII.


Brown, D.G. *Action.* University of Toronto Press, 1968. On facts being reasons see Secs. 1.4, 1.8, 3.9.


Geach, P.T. "Dr. Kenny on Practical Inference." *Analysis*, XXVI (1966), 76-79.


Gibson, Q. *The Logic of Social Enquiry*, New York, 1960. Ch. IV.


Ross, W.D. *Aristotle.* New York, 1959. Ch. VII.

Sachs, D. "On Mr. Baier's 'Good Reasons'." *Philosophical Studies,* IV (1953), 65-69.


Toulmin, S. *An Examination of the Place of Reason in Ethics.* Cambridge, 1950.


The works listed below contain several of the papers referred to in Sections I – VI of the bibliography, often together with useful introductions to the problems. Most of them also contain readings on topics bordering on those of headings I – VI, such as causation, explanation, free-will, the mind-body problem, etc.


