RATIONAL AGENCY

By

PETER G. CAMPBELL

B.A., The University of British Columbia, 1983

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS

in

THE FACULTY OF GRADUATE STUDIES

(Department of Philosophy)

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

April, 1988

© Peter G. Campbell, 1988
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 or her 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
1956 Main Mall
Vancouver, Canada
V6T 1Y3

Date April 15, 1988
ABSTRACT

It is claimed that action discourse provides us with a criterion of adequacy for a theory of action; that with action discourse we have a family of concepts which a theory of action must accommodate. After an exegesis of Davidson's essay "Agency", it is argued that his semantics of action is incompatible with our concepts of motivation and responsibility for action and of attributions of action and agency, and must, therefore, be rejected. A theory of rational agency is presented within which are to be found accounts of intention, coming to intend, intentional action, and an alternative semantics of action which connects the action essentially to agency. The theory of rational agency is then used to illuminate the concepts of trying, compulsion, autonomy and involuntariness, mistake, accident, and the so-called active-passive distinction.
# TABLE OF CONTENTS

ABSTRACT ........................................................................................................ ii
TABLE OF CONTENTS .................................................................................. iii
LIST OF FIGURES ........................................................................................ v
ACKNOWLEDGEMENT .................................................................................. vi

INTRODUCTION ............................................................................................. 1

## CHAPTER I

THE SEMANTICS OF ACTION ......................................................................... 4
A. DAVIDSON'S "AGENCY" .......................................................................... 5
   1. The Criterion of Action and the Expression of Agency .................... 5
   2. Agency. The Role of Causation in Action, Part I .............................. 10
   3. The Role of Causation in Action, Part II ........................................ 17
   4. Summary ......................................................................................... 26
B. DAVIDSON'S SEMANTICS REJECTED .................................................. 26
C. SOME IMPLICATIONS FOR A SEMANTICS OF ACTION .................... 39

## CHAPTER II

RATIONAL AGENCY ...................................................................................... 43
A. PROPERTIES OF INTENTION ................................................................. 44
   1. Davidson on Intention ..................................................................... 44
   2. Bratman on Intention ..................................................................... 49
B. A THEORY OF RATIONAL AGENCY ..................................................... 52
C. INTENTIONAL ACTIONS ...................................................................... 71
CHAPTER III
APPLICATIONS

A. AGAINST THE SIMPLE VIEW ........................................ 77

B. TRYING ........................................................................ 84

C. NON-STANDARD ACTIONS AND EXCUSES ......................... 90
   1. Mistake .................................................................... 90
   2. Accident .................................................................... 93
   3. Compulsion ............................................................... 95
   4. Involuntariness .......................................................... 96

D. THE ACTIVE - PASSIVE DISTINCTION .............................. 98
   1. Intentional Letting Die ................................................. 100
   2. Negligent Letting Die .................................................. 102

BIBLIOGRAPHY ................................................................. 106
<table>
<thead>
<tr>
<th>Fig.</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A MODEL OF RATIONAL AGENCY: RAAG</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>RAAG PLAYING THE VIDEO GAME</td>
<td>53</td>
</tr>
<tr>
<td>3</td>
<td>RAAG MAKES A MISTAKE</td>
<td>58</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td>58</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>60</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>61</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td>62</td>
</tr>
<tr>
<td>9</td>
<td></td>
<td>64</td>
</tr>
<tr>
<td>10</td>
<td></td>
<td>64</td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>65</td>
</tr>
<tr>
<td>12</td>
<td>A MODEL OF RATIONAL AGENCY: RAAG</td>
<td>66</td>
</tr>
<tr>
<td>13</td>
<td>RAAG PLAYING THE VIDEO GAME</td>
<td>82</td>
</tr>
<tr>
<td>14</td>
<td>RAAG MAKES A MISTAKE</td>
<td>91</td>
</tr>
<tr>
<td>15</td>
<td>RAAG ACCIDENTALLY SHOOTS HIS NEIGHBOUR'S DONKEY</td>
<td>95</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENT

I would like to thank Professor S.C. Coval for two years of work in action theory. This thesis is the product of our close collaboration which began with Professor Coval's action theory seminar in 1986, and which continues with further research on the problems of action, and applications of our theory to artificial intelligence. We are working with Professor J.C. Smith and Doug Arnold of the U.B.C. Law School to construct a computer model of a rational agent. Chapter II of this thesis will form part of a paper co-authored by Professor Coval and I in which we discuss our work on the computer model.

I would like to express my appreciation both to the U.B.C. Philosophy Department for financial support in the form of teaching assistantships, and to the University of British Columbia for a Graduate Summer Fellowship, 1987.
INTRODUCTION

This essay is concerned with some of the main problems in the philosophy of action. These include the semantic issues of the nature and identity of actions and action descriptions, action explanations, both causal and teleological, and assessment of agents. I take it as a given that there is mental causation. With mental causation we accept the notion of mental representation with cognitive content and causal efficacy. I argue that there is a set of properties which the agency state must have in order for there to be action. However, even though the nature of mental states and events is a concern of the philosophy of mind, I claim that nothing I have to say about rational agency decides, in any way, any of the traditional problems in the philosophy of mind. That is, the theory of action which I present in Chapter II does not beg any questions for or against dualist, materialist or functionalist accounts of mind. It is rather that with action, we accept a set of concepts which any viable theory of mind must accommodate.

Our pre-theoretical action discourse presents us with a criterion of adequacy for a theory of action. With action discourse, we have a set of concepts and distinctions which we employ in complex ways. We are at home with causal and
teleological explanations of actions. We assess agents in order to determine degrees of responsibility, and we wield excuses and defenses. It is a deep-seated fact of our conceptual scheme that purposeful behaviour is amenable to this kind of scrutiny. Therefore, a theory of action ought to accommodate and clarify these concepts which are the given of action theory. And without good reason to do otherwise, we ought to minimize the extent to which the shape of a theory disrupts our pre-theoretical concepts. What counts as an action must pay off to our concern for classifying events and agents in relation to our concern with responsibility and explanation of action. This is not to say that we ought to let pragmatic considerations such as our concern with holding persons responsible for events determine our scheme of classifying events as intentional actions or otherwise. That is, we ought not allow the inference from the claim that you are responsible for some event A to the claim that you intended A, nor to the claim that you desired A, and therefore not to the claim that you did A intentionally. Nevertheless, any theory of action whose semantics or other theoretical aspects is inhospitable to action discourse violates the criterion of adequacy. Unless there is some good reason to the contrary, such a theory ought, therefore, to be rejected.

In Chapter I, I argue that Donald Davidson's semantics of action is actually incompatible with our family of action concepts. After an exegesis of his theory, I argue that it cannot accommodate the notions of responsibility and motivation,
nor the classifications mentioned above, and therefore cannot accommodate agency.

In Chapter II, I present a theory of rational agency within which are to be found accounts of intending and acting intentionally, and an alternative semantics of action which connects the action essentially to agency. The theory requires that we accept such "metaphysically suspect" items as intensional objects, but this, I argue, we already accept with action.

In Chapter III, I apply the theory of rational agency, first to an objection of Michael Bratman's, and then to some of the long-standing problems in action theory, namely trying, compulsion, autonomy and involuntariness, mistake, accident, and to the so-called active-passive distinction.
CHAPTER I

THE SEMANTICS OF ACTION

I begin with an exegesis of Davidson's "Agency"\(^1\), in which he develops his semantics of action and analyses the notion of agency. A close look at "Agency" will reveal some of the main questions in action theory and the problems which Davidson reveals in the positions of others. Furthermore, no other action theorist is as powerful on, for example, the issue of the nature and identity of actions. However, Davidson is the principal proponent of the view of action which I oppose. For Davidson, actions are events. Events are metaphysically respectable items which, like objects, may be described in various ways.\(^2\) Therefore, actions are extensionally conceived events. If so, I argue, the semantics of action is cut off from intention with the consequence that actions cannot support the normal inferences between what is done, and for example, motivation and responsibility. Following the exegesis, I begin to work my way out of the grip of the Davidsonian semantics with several arguments against his extensionality. In Chapter II, I

---


\(^2\)Ibid., 105-203.
present an alternative semantics of action within a theory of rational agency.

A. DAVIDSON'S "AGENCY"

"Agency" is concerned with the semantics of action. Davidson's stated aim is to find an analysis of agency. This he takes as the task of giving an analysis of the relation that holds between an agent and his actions, and an analysis of the relata which stand in this relation. There are, roughly, three sections to Davidson's essay: the first addresses the nature and identity of the second term of the relation - the action. The second and third sections deal mainly with the relation itself. By his own admission Davidson does not give us an analysis of agency, but does show what it cannot be by revealing some mistakes of other action theorists.

1. The Criterion of Action and the Expression of Agency

The first, and for our purposes most important section of "Agency" addresses the question "What is an action?" Davidson assumes that "there is a fairly definite subclass of events which are actions." This gives us the genus of actions: they are events. The problem is to distinguish the action subclass. Davidson's first candidate for the distinguishing mark of actions is that of being intentional. Since there are actions

3 Ibid., xiii.
4 Ibid., 43-47.
5 Ibid., 44.
which cannot be anything but intentional, e.g., asserting, cheating and lying,\(^6\) intention implies agency. But, the converse does not hold: that what you did was an action does not imply that you did it intentionally.

If...I intentionally spill the contents of my cup, mistakenly thinking it is tea when it is coffee, then spilling the coffee is something I do, it is an action of mine, though I do not do it intentionally.\(^7\)

Therefore, since agency does not imply intention, an event's being intentional is not sufficient as a mark of action.

Davidson's treatment of mistakes is telling. Mistakes cannot be done intentionally. Making a mistake, he says, is failing to do what one intends, is "doing something with the intention of achieving a result that is not forthcoming."\(^8\) Mistakes are, therefore, unintentional. Yet for Davidson they are actions.\(^9\) For example, a misreading is a reading, a misinterpreting is an interpreting. So, making a mistake is doing something else intentionally:\(^10\) an unintentional misreading is an intentional reading; an unintentional misinterpreting is an intentional interpreting. And, mistakenly

\(^6\)And not acting!

\(^7\)Davidson, 45.

\(^8\)Ibid., 46.

\(^9\)This will be a point of dispute between us.

\(^10\)As will be apparent, Davidson does not really mean "something else".
spilling the coffee is intentionally spilling the contents of the cup.

What is common to all of the above cases of agency is that the agent is doing something intentionally. What cancels agency, for Davidson, is when the event in question was caused externally: "I am the agent if I spill the coffee meaning to spill the tea, but not if you jiggle my hand."\(^\text{11}\) In the former case I am doing something intentionally, in the latter I am not. (Of course, I am doing several things intentionally in the latter case, including holding the cup so as not to spill its contents. What Davidson means is that the bodily event which causes the event of the contents spilling cannot be described so as to make the spilling intentional for me.) So, my spilling the contents of the cup was intentional. This very same act can be redescribed as my spilling the coffee.

On the basis of this analysis, Davidson offers the following causal criterion of agency, and so of action:

**Criterion I**

A man is the agent of an act if what he does can be described under an aspect that makes it intentional.\(^\text{12}\)

That is, doing something which can be described under an aspect that makes it intentional is sufficient for agency or action.

\(^{11}\text{Davidson, 46.}\)

\(^{12}\text{Ibid.}\)
What makes this possible is that attributions of intention are intensional: co-designating singular terms which refer to what one does intentionally cannot be substituted, salva veritate. If this were not so, then from

1) Hamlet intentionally kills the man behind the arras.
and 2) The killing of the man is identical to the killing of Polonius.

we could infer

3) Hamlet intentionally kills Polonius.

which is false. Whether or not a doing is an intentional doing depends on how it is described. I spill the contents of the cup. This action can be redescribed as my spilling the coffee. Under the first description, my action was intentional. Under the second description, the very same act was unintentional. This doesn't mean, of course, that the event which is the action both had and lacked a certain property - that of being intentional. It means that the event, the agent, and a certain description have a relation that does not obtain between the same event, the agent, and a different description.\textsuperscript{13} So, while there is a class of events which are actions, there is neither a class of intentional actions, nor a class of unintentional actions: what I did will be intentional under one description, and not under another.

The criterion of agency and action, put in terms of sentences and descriptions of actions is:

\textsuperscript{13}Ibid., 195.
Criterion II.

A person is the agent of an event if and only if there is a description of what he did that makes true a sentence that says he did it intentionally.\textsuperscript{14}

That is, it is sufficient and necessary for agency and action that there be a \textit{single} description of what a person did that makes true a sentence that says he did it intentionally.

While attributions of intention are intensional, attributions of agency or action are purely extensional: "the expression of agency is itself purely extensional."\textsuperscript{15} Singular terms which refer to what an agent did are substitutable, \textit{salva veritate}. As we have seen from Davidson's examples, being an action is a trait which particular events have independently of how they are described: my spilling the contents of the cup is identical to my spilling the coffee; Hamlet's killing the man behind the arras is identical to his killing Polonius; an officer's sinking of that ship, which he mistakenly thinks is the Tirpitz, is identical to his sinking the Bismark.

"The relation that holds between a person and an event, when the event is an action performed by the person, holds regardless of how the terms are described."\textsuperscript{16}

\begin{flushleft}
\textsuperscript{14}Ibid., 46.
\textsuperscript{15}Ibid., 47.
\textsuperscript{16}Ibid.
\end{flushleft}
Therefore, we can say that a person does as agent whatever he does intentionally under some description.

This is Davidson's semantics of action. Actions are extensional events which satisfy his criterion of agency and action.

2. Agency. The Role of Causation in Action, Part I

The second main section of "Agency" is a search for an analysis of agency that does not appeal to that mysterious and, as yet, unanalyzed notion of intention. What Davidson wants is an analysis of agency, of the relation that holds between an agent and his actions. One way to arrive at an analysis of a relation is through an analysis of the relata - of the things which can stand in the relation. We know now that the second term, for Davidson, is an extensional event. Which events are just those which satisfy the criterion. The criterion of agency and action appeals to the notion of intention. So, if we had an analysis of intention (a good place to begin an analysis of agency) we could ask for an analysis of the relation that holds between intentions and their corresponding actions. But Davidson had no analysis of intention, discounting the "old" one from "Actions, Reasons and Causes", where the intention with which an action is done is analyzed in terms of the desires and beliefs which rationalize the action. Therefore, the question

\[
^{17} \text{Ibid., 47-55.}
\]

\[
^{18} \text{Ibid., 3-19.}
\]
Davidson asks is "what can we say, without an analysis of intentions, about the relation between an agent and his action that will elucidate doing, acting, and so agency?" His suggestion is that we pick up the notion of the expression of agency.

With the extensionality of the expression of agency we have the idea that if an event is an action, then the relation that holds between the agent and the event holds regardless of how the terms are described. Since intentions are intensional and agency not, it looks as though the concept of agency is simpler than that of intention. And, the relation itself looks, to Davidson, a lot like ordinary event causality. This being the case, an examination of the role of causation in action ought to clarify the concept of agency.

a. There are Primitive Actions

The relation of event causality is the relation that holds between actions and their upshots, and according to Davidson, it is this which allows the attribution of the upshots to the agent: it is because an action causes an upshot that the upshot is attributable to the agent. So some events which we attribute to an agent are events related causally to some other event of which he is the agent. However, not every action attributed to an agent is so attributed because it is caused by another act of the agent. Some actions must be primitive actions — that is,

---

19Ibid., 47.
not caused by some causally prior act of the agent. Otherwise, agency could never get started. Therefore, event causality cannot be the relation which primitive actions bear to the agent, and so event causality cannot support the first attribution of agency.

b. What are Primitive Actions?

Davidson identifies primitive actions as bodily movements. (or, perhaps better, bodily movings, as opposed to mere movements of the body). He considers two main objections, the first of which argues that bodily movements are neither primitive nor actions, the second of which argues that primitive actions involve more than body movement.

It may be argued that in order to move my finger I must first do something that causes my finger to move, namely, contract the appropriate muscles, and, doing this requires that I first make certain brain events occur. Since neither of these causally prior events are bodily movements, bodily movements are not primitive actions.

Accepting that movements of the body are caused by contractions of certain muscles, which in turn are caused by certain brain events, Davidson's response is: "Doing something that causes my finger to move...is moving my finger".\(^\text{20}\) The trick is turned, here, by distinguishing the event of the

\(^{20}\text{Ibid.}, 49-50.\)
movement of the finger from the action of moving the finger: moving the finger - that primitive action/bodily movement is doing whatever it takes to cause the event of the finger movement.

Chisholm\(^{21}\) has two objections, one addressing Davidson's particular point here, the other addressing extensionality, and both based on the claim that knowing that one is doing A is a necessary condition for one's doing A. If so, making cerebral events and muscle contractions occur, when these are unknown, are not actions. And, since consequences of actions are often unknown, there are effects of actions which are not actions (contra Davidson).\(^{22}\)

Davidson has a counter example\(^{23}\) to the necessity claim: Suppose that a man intends to make 10 carbon copies as he writes but does not know or believe with any confidence that he is succeeding. If he does, Davidson claims, he does so intentionally, and without knowing that he is. Therefore, knowing that one is A-ing is not a necessary condition of one's A-ing.\(^{24}\)


\(^{22}\)My arguments against Davidson's extensionality will support Chisholm's claim.

\(^{23}\)Davidson, 50, 92.

\(^{24}\)For a detailed discussion of this matter, see Chapter III, Trying.
Knowing that one is A-ing is not a necessary condition for A-ing, for Davidson. Nevertheless, that his criterion of action must be satisfied entails that one must know whatever one is doing under some description or other. Suppose O moves his finger intentionally. If so, O intends to do with his body whatever is needed in order to produce the event of the finger movement, where the "whatever needed" is the cerebral events and muscle contractions. Therefore, in intending to do whatever is needed, O intends and knows about the cerebral and mental events under the description "whatever it is I do when I move my finger".

The second main objection to Davidson's claim that primitive actions are bodily movements is the claim that some primitive actions include more than bodily movements. For example, tying one's shoes includes the finger movement and the movement of the laces. The question is, can these be separated into two events, the first of which is the primitive action? Since events are distinguished by their space-time locations, the separation here is not the problem. The problem is whether there is a description of what is done that refers to the event of the body moving and not to the shoelaces. Davidson's response is the same as he made to the last objection. The description "I move my body in just the way required to tie my shoelaces" satisfies the criterion, is known to the agent, and refers to the bodily event. It is a description of an event with the right effects, but no less a description of a bodily event which is a primitive action.
c. Agent Causality

Returning to the question of the relation between an agent and his action, we can now take, as the second term, primitive actions. As we have seen, Davidson argues that while ordinary causation can explain how agency spreads from primitive actions to actions described in further ways, it cannot explicate the relation between the agent and his primitive actions. The question now is whether there is a *sui generis agent causality*\(^{25}\) which relates agents and their actions, and which will, therefore, explicate the relation between an agent and his primitive actions. Another possibility which Davidson doesn't consider is that there is a term which, being unique as a causal term (desire), helps explicate that relation.

Davidson's treatment of agent causality is important for my thesis since it reveals the ground upon which our battle will be fought. Agent causality is supposed to be that which relates agents to events and by virtue of which these events are actions. Davidson poses the following dilemma for friends of agent causality. Either the causing of a primitive action is an event discrete from the primitive action, or it is not. On the first horn, agent causality introduces an event separate from, and prior to, the primitive action. There are two possibilities. First, this prior event is an action. If so, then contrary to the assumption, the primitive action is not

\(^{25}\)From Thalberg.
primitive. And, if agent causality is to serve its function, then this prior event which is an action must be related to another event, separate from and prior to it. And we have a regress. Therefore, agent causality could never do its work, and agents could never act.

The second possibility for the first horn of the dilemma is that the separate and prior event is not an action. But if so, it is an event which, because of agent causality, is an agent-causing which is not a doing. Therefore, agent causality is not doing the job it was introduced for, since agent-causings are not actions. Davidson concludes that agent causality is therefore superfluous.

It is important to notice that Davidson's own position is one in which primitive actions are events which are causally related to separate and prior brain events which are not actions. Since the causally prior events are not actions, and given the nature of brain events, for Davidson, he can stick with good old ordinary event causality, as the relation which they bear to primitive actions.

On the second horn of the dilemma, agent causality does not introduce an event separate from and prior to the primitive action. If so, the agent is not related by an event of agent causality to his primitive actions. Thus the notion of agent causality would play no role in agency, and saying that one is related by agent causality to his actions would be saying no
more, and be nor more explanatory, than saying he acted. Again, agent causality is superfluous.

The problem Davidson must face is that if the agent is not related by some notion of agent causality to his actions, then we are without grounds to support the claim that they are actions. If the agent is merely causally related to his actions, and since we as agents are related causally to any number of events which are not actions, then there would seem to be nothing to support the distinction between actions and events which are not actions.

3. The Role of Causation in Action, Part II
   a. The Accordion Effect

So far, Davidson has argued that causation is central to the concept of agency, that it is garden-variety event causality, and it concerns the effect of actions and not their causes. The feature of action discourse called the accordion effect\(^\text{26}\) demonstrates this role of causation in action. Suppose the following:

1. \(0\) moves his finger intentionally.
2. \(0\) causes the switch to be flipped.
3. \(0\) causes the light to come on.
4. \(0\) causes the room to be illuminated.
5. \(0\) causes the prowler to be alerted.

\(^{26}\text{From Feinberg.}\)
Since O's moving of his finger is intentional, it qualifies as an action by Davidson's criterion. But, whether or not O intends any of the consequences 2-5, Davidson says that 1-5 entail:

6. O flipped the switch.
7. O turned on the light.
8. O illuminated the room.
9. O alerted the prowler.

All of 6 - 9, as well as 1, are action descriptions, and are attributable to O because of the accordion effect, which reveals that causation transfers agency from actions to the effects of actions, whether or not the agent intends those effects:

...once he has done one thing (move a finger), each consequence presents us with a deed; an agent causes what his actions cause.27

The accordion effect demonstrates that causation is what transports agency, though not intention, and intention is irrelevant as to how the accordion is played.

27Davidson, 53. Except for the difficulties which Davidson footnotes, he would like to be able to say that an agent does what his actions cause. It is this sort of feature of Davidson's theory that I object to when I object to his extensionality.
The accordion effect works because of causation between events. But it is a feature of action discourse only. Therefore, the relation it reveals is not ordinary event causality. Causation is a transitive relation between events: if event A causes event B, and B causes event C, then A causes C. But it is not this property of causation that powers the accordion:

(i) Suppose the bat struck the ball which broke the window. If so, we can say that the movement of the bat broke the window, but not that the bat broke the window.

(ii) Suppose the cue struck the cue ball which moved the 8-ball. We can say that the movement of the cue caused the movement of the 8-ball, but not that the cue moved the 8-ball.

(iii) Suppose that 0 flipped the switch which caused the light to go on. We can say both that the movement of his finger caused the light to go on and, that 0 turned on the light.

The relation between events in (i), (ii) and (iii) is event causality. The relation between 0 and his action of turning on the light is not. Just what the relation is between an agent and his non-primitive actions is addressed in the next section on "Agency".
b. Doing With and Doing By

The third main section of "Agency"²⁸ is concerned with the relation between an agent and his non-primitive actions, which is the relation of "doing with" and "doing by" in the following:

A. Brutus killed Caesar by stabbing him.
   (2 non-primitive actions)

B. Jones killed Smith by startling Smith by opening the door by turning the key.
   (All non-primitive actions)

C. With one movement of his hand, Jones did all of the things in B. (A primitive action and several non-primitive actions)

Davidson has argued that primitive actions cause their consequences, and that this is ordinary event causality. If some event A is a primitive action (that is, an event caused in the right way) and if events are related causally to their effects, then primitive actions are related causally to their effects. Similarly, we can see that non-primitive actions cause their consequences where again this is event causality: the stabbing of Caesar by Brutus caused the death of Caesar.

²⁸Davidson, 55-61.
However, we as yet have no analysis of the relation that holds between and among primitive actions and non-primitive actions. First, we want to know how, e.g., the primitive action of moving one's hands is related to the non-primitive action of tying one's shoes. Secondly, we want to know how non-primitive actions are related to certain correlative events: e.g., how is the non-primitive action of tying one's shoes related to the event of one's laces being tied? Or, how is the non-primitive action of Brutus killing Caesar related to the death of Caesar? Finally, we need to know what the relation is between non-primitive actions: e.g., what is the relation between the two non-primitive actions of Brutus stabbing Caesar and Brutus killing Caesar? The obvious candidate is causation. But Davidson shows why this cannot be.

c. Feinberg, Austin and Danto.

On Feinberg's version of the accordion effect, the squeezing down and puffing out of an accordion is done over events, so that an action can be stretched out to include an effect. But, as Davidson notes, if the squeezing and the stretching changes the time span of the event, then it cannot be the same event, and so cannot be the same action. Therefore, this is not a correct analysis of the accordion effect. Feinberg also has a distinction between simple and causally complex acts. A simple act is one which requires us to do nothing else. A causally complex act is one which requires us to do something first as a means. For example, in order to open the door I must first do something else which will cause the
door to open. Therefore, on Feinberg's view, there are causally connected sequences of acts.

J.L. Austin and Danto have similar views. For Austin

...a single term descriptive of what he did may be made to cover either a smaller or a larger stretch of events, those excluded by the narrower description being then called "consequences" or "results" or "effects" or the like of his act.29

Therefore, actions can cause actions, for Austin. Danto makes the distinction between basic and non-basic actions, where the distinction is marked causally: A basic action, such as moving a hand, causes the non-basic action of moving the stone.30

Davidson argues that these views are mistaken; that actions are not causally related. While it is true that primitive actions cause their consequences, so that when I, for example, close the door, some action of mine causes the event of the door closing, it is false that some action of mine causes my action of closing the door. If not false, then in order to close the door - that action - I would have to do two things. But it seems clearly false that when I close the door by moving my hand I do two things. And, if there are not two things, then they cannot be related, causally or otherwise.


The heart of Davidson's position here is that a primitive action is not numerically distinct from its corresponding non-primitive action. If so, then there are not two events to be causally related. Otherwise, if there were two actions causally related, we would have to say that, for example, the queen's moving her hand caused the queen's killing the king, and that the queen caused the queen to kill the king. Suppose this were the case. We still want an analysis of the action of the killing. Here, the killing is the queen's action of moving her hand in just that way in those circumstances. But, this was sufficient in those circumstances, to kill the king. If so, there was no room for any further action of the queen:

Is it not absurd to suppose that, after the queen has moved her hand in such a way as to cause the king's death, any deed remains for her to do or to complete? She has done her work; it only remains for the poison to do its.\(^{31}\)

The action of the queen's killing the king cannot be analyzed as the action of the queen's moving her hand causing the action of the queen's killing the king. Actions are not related causally.

Davidson has argued that it is a mistake to suppose that an action begins with a primitive action but ends later. So, the killing of the king does not begin with the queen's primitive action of moving her hand and end with one of its consequences,

\(^{31}\)Davidson, 57-58.
the king's death. It is true that the hand moving caused the death, but "the queen moving her hand in that way, in those circumstances" and "the queen doing something that caused the death of the king" are two descriptions of the same event. And doing something that causes a death is causing a death. Causing a death is, for Davidson, a killing. Therefore the killing of the king is the queen's moving her hand in that way, in those circumstances, with that effect.

d. Davidson's Accordion

Davidson's diagnosis of the difficulty which others have had with action is the tendency to mistake features of a description of an event with features of the event described. An event can be described in terms of its causes or of its effects. Here are two definite descriptions of some event \( E_2 \):

(1) the event caused by event \( E_1 \)

(2) the event which caused the event \( E_3 \)

It is a mistake to suppose that the consequence, \( E_3 \), is included in the definite description in (2). What changes between (1) and (2) is the description, not the event described: (1) and (2) are co-referential definite descriptions of one event, \( E_2 \). As applied to action discourse, we can describe a single action in terms of its causes or of its effects. What gets described, if we do, is an event which is the action. So, for Davidson, the accordion which endures the squeezing and the stretching of the accordion effect is a single event which is an action. What
changes is the description of the action. Therefore, all of the following correspond to a single descriptum, which is the action:

(i) "The queen moved her hand."
(ii) "...thus causing the vial of poison to empty into the king's ear."
(iii) "...thus causing the poison to enter the body of the king."
(iv) "...thus causing the king to die."
(v) "The queen moved her hand thus causing the king to die."
(vi) "The queen killed the king."
(vii) "The queen emptied the vial into the king's ear."
(viii) "The queen killed the king by pouring poison in his ear."
(ix) "The queen poured poison in the king's ear thus causing his death."^{32}

The single descriptum is the bodily event which is the primitive action - the queen's moving her hand. Therefore, Davidson says, somewhat paradoxically, that there are only primitive actions:

\[
\text{We never do more than move our bodies; the rest is up to nature.}^{33}
\]

\[^{32}\text{On the account of action which I develop (Chap. II), neither (i), (i) + (ii), (i) + (iii), nor (vii) count as descriptions of the queen's action.}\]

\[^{33}\text{Davidson, 59.}\]
The concept of being primitive is, like the concept of being intentional, intensional: under some description, namely (i) above, what the queen did was primitive; under another, say (vi), it was not. Likewise, the concept of being non-primitive is intensional: under some description, e.g., (vii), what the queen did was non-primitive; under another, e.g., (i), it was not.

4. Summary

The positive parts of Davidson's theory of action which he develops in "Agency" are, then, as follows:

1. The Criterion: Actions are events which are intentional under at least one description.

2. The role which causation plays in action is the role it plays everywhere; namely, it is the relation between causes and effects, and it is the relation of causation which permits re-description of actions in terms of their causes and effects.

3. Primitive actions are all the actions there are: "There are no further actions, only further descriptions."34 And, non-primitive actions are all the actions there are.

4. Agency: The relation which holds between a person and an event which is his action is independent of how the terms of the relation are described.

B. DAVIDSON'S SEMANTICS REJECTED

In this section I argue that Davidson's criterion of action and agency cannot work together with his semantics within a theory of action. I then suggest an alternative, leaving its

34Davidson, 61.
full development to the next chapter. First, it will pay to restate his notions of criterion and expression.

Davidson's criterion of action, which tells us whether something was done, and so whether agency is on hand, is that an event is an action if and only if it is effected through agency. This is a causal criterion, the causal term being an intentional entity or an agent. Since we want to distinguish arm raisings from arm-risings, it is necessary that events which are actions be causally related in the right way to an agent, where "the right way" is cashed in terms of intention: the agent must have intended the event under one true description. In other words, an event is an action if and only if it is in the right relation to the right states of an agent.

Davidson's expression of agency, on the other hand, addresses another aspect of the relation between action and the agent: it tells us what was done, and so identifies what was caused by agency. The expression of agency addresses the semantics of action. Davidson believes that extensionality reigns among actions, that what you do when you act is your action under any of its true descriptions. If you spill the contents of your cup mistaking it for tea when it is coffee, then you spill the coffee. If the officer sinks the ship mistaking it for the Tirpitz when it is the Bismark, then he sinks the Bismark. Because the expression of agency is purely extensional, descriptions of what was done function extensionally. All non-equivalent but co-designative
descriptions of an event which satisfies the criterion pick out the same action. Therefore, as long as an event is an action—so long, that is, as it meets the criterion—it can be ascribed to the agent as his action under any description true of it. The concept of an action is an event extensionally conceived, for Davidson.

We can begin to see that the criterion and the expression cannot work together if we consider the following two points. First, we need a conception of action, and so of agency, which will allow us to distinguish cases in which full agency has been at work, from those cases in which partial or diminished agency has functioned. This distinction is employed in a rich way in action discourse. It is that which grounds the distinction between full and partial responsibility, which in turn accommodates the workings of excuses and defenses. The full/partial agency distinction, and so excuses, requires that we recognize (is the recognition?) that acting standardly differs from acting non-standardly.35 A criterion of action and agency ought, therefore, to mark this distinction. Davidson's criterion cannot. Indeed, any criterion based, as his is, on ordinary event causality where the terms of the relation are treated extensionally is, for this reason, incapable of marking the above distinction. For Davidson, the acting which occurs in standard actions does not differ from the acting in non-standard

actions, nor would his criterion allow the distinction. The fact of it being an action is based only on the causal relation.

Secondly, the distinction between full and partial agency implies a corresponding distinction among attributions of actions: there are standard actions and non-standard actions; those of the latter type being deviants of the former. This distinction is what is marked by "intentional" and "unintentional" in action attributions. There are intentional actions and unintentional actions, like mistakes, and accidents.\textsuperscript{36} Conceptually we cannot make sense of the non-standard without a theory of the standard.\textsuperscript{37} The extensionality of Davidson's expression of agency will not permit the distinction since what gets attributed to an agent as his action is, in standard and non-standard cases, the extensionally conceived event.

I have three arguments to support these claims against extensionality.

1. The Argument from Motivation

An adequate theory of action should leave undisturbed the normal inferences between action - what was done - and

\textsuperscript{36}The term "unintentional action" is highly misleading. If actions are necessarily intentional, then with failure of intentionality we have something other than an action: we have an unintentional event.

motivation. To describe an event as an action of an agent is to impute a motive to that agent, the content of which is embedded in the description of the action attributed to him. That an action was a standard action implies, among other things, that it was successful. To be successful is for the action to satisfy the motivation which prompted it - where this implies that the motivation was the cause of that which satisfied it. What satisfies the desires which constitute the motivation will tell us what the desires were for and will therefore, in standard cases, be a truth-condition on what was intended.

Satisfaction is an intensional notion, and so will occur only if some event is cognized, and cognized under the description under which it was desired. For example, the desire that the government be embarrassed will be satisfied by an event cognized under that description, and not by the same event cognized under the description "the government voluntarily changing its policy" or "the government recessing for holidays". Therefore, what you desire when you desire the government's embarrassment is not an event under any of its true descriptions. What you intend, when you intend to embarrass the government, is not to produce an event under any of its true descriptions, nor all of its true descriptions. What you intend, when you intend to produce that which will satisfy the motivating desire, is an aspect of an event. If so, and if we are to satisfy the desideratum that what you intend to do and what you do intentionally are tightly linked (i.e., if we are to have a unitary theory of intending and doing intentionally),
then what you do as the result of a successful intention to produce an aspect of an event is to produce an event-under-a-description, where the description is the one embedded in the propositional content of the desire(s) which motivated the doing.\textsuperscript{38}

Davidson's semantics of action, of what is done, is inadequate in not providing for the essential motivational ingredient of action. Since motivation is an intensional notion, the normal inferences between action and motivation cannot be supported by an extensional semantics of action. If the expression of agency is purely extensional, then what an agent does is what he intends, whether or not he intends it under the description contained in the ascription. Therefore, the connection between the content of the desires and intention in the causal ancestry of the action and the action, is severed. Thus even Davidson's criterion of action is so severed from the semantics that it cannot play a motivational role. Under the limits of a merely causal criterion of action, which is what Davidson's becomes, since it is separated from the expression of agency, there are no means to explain why or to test whether the resultant event was intended. These questions make sense only when they essentially implicate some description of the event. "What did you intend?" and "Why did you do that?" cannot be answered extensionally. If the right inferential relations

\textsuperscript{38}Bratman has an objection to "tight-fit" which I address in the sequel, where the relation of desire and belief to intention is developed.
implied by motivation are to be retained, neither can the question "What was done?" be answered extensionally.

2. The Argument from Responsibility

To describe an event as an action of an agent is to ascribe responsibility for that event to the agent. This gives us one right answer to the question about what we are responsible for. We can also be held responsible for events which are not actions. The charge of negligent commission ascribes to the agent responsibility for a consequence of an action, typically when the consequence is negative, and only when one should not have caused the consequence. Negligence implies a standard of care with respect to action according to which one should have foreseen the likelihood of the negative outcome of an action, and should have, for this reason, refrained from so acting. Where one fails to meet the standard of care and thus fails to foresee the negative outcome, or acts despite the belief that there would be negative consequences, you are said to have acted negligently and are held responsible for the negative consequence. The charge of negligent omission ascribes to one responsibility for an event when one should have acted, counterfactually, to prevent some, typically negative, outcome. Where the charge of negligent omission is warranted, one is held responsible for a consequence, not of one's action, but of some event which you failed to prevent, and should have prevented.\footnote{For a detailed discussion of the relation between negligence and responsibility, see Chap. III, The Active-Passive Distinction.}
Therefore, we are responsible for our actions, and can be held responsible for the consequences of our actions and inactions. This is to say that responsibility can carry further down the causal chain of events than does agency.

The degree to which one is held responsible for an action varies with the conditions of acting. One can be held fully or only partially responsible for some event. The distinction between full and partial responsibility goes hand in hand with the distinction between full and partial agency. Responsibility for an event diminishes as agency of the event diminishes. It is the job of excuses and legal defenses to reduce the degree to which one is held responsible for some event, and the reduction, where justified, is justified because agency is at work in the causal genesis of the event in some diminished way.

An adequate theory of action should leave undisturbed the normal inferences not only between motivation and action but also between action and what one is held responsible for. Clearly, one is held responsible for one's actions. On Davidson's extensional account of action I am held responsible for what I, as agent, cause however described so long as there is a true description of the event under which it was intentional for me. But this prevents me from employing as a defence against the ascription of responsibility, the fact that I did not intend that event under the description contained in
the ascription. I am so prevented because it is that extensional event, however described, which is my action. If, therefore, Davidson is right, and actions are the sheer events we cause as agents, and if we are responsible for our actions, then the description under which I intended the event should be irrelevant to how we gauge responsibility. But, it is not irrelevant to responsibility. The significance for responsibility of the description of events which figure in one's intention is that it is just this which allows excuses to work in determining degrees of responsibility. Whenever an event which is caused through my agency is ascribed to me under some description other than that which figured or should have figured in my intention, then an excuse applies and I am not responsible for that event under that description precisely for this reason. Just such a move is blocked if actions are conceived of extensionally.

Consider the following argument. We are not responsible for our mistaken actions. If I knowingly and intentionally, shake the hand of the man I truly believe to be the bank manager, then shaking his hand is an action of mine, and I am therefore, responsible for it. Suppose that he is, unbeknownst to me, the bank robber. On Davidson's view, shaking the bank manager's hand is the very same action as shaking the bank robber's hand. If we are responsible for our actions, then I ought to be responsible for that very same action, shaking the

40 The difference between intending and not intending is like the difference between desiring and not desiring.
hand of the bank robber. But I am not. Therefore, I am not responsible for an event under a description under which it was not intended.

That we employ excuses and defenses as we do suggests that one is responsible, not for sheer events, but only for events under the description under which it occurred in one's successful intention. And, if we are responsible for our actions, then, by substitution, an action is an event essentially under the description which figured in its cause. To call an event an action is to so describe it. It would appear that Davidson is without a theory of excuses and so of how responsibility is gauged, and is without an account of how desires and beliefs function in agency so long as he holds on to his purely extensional account of agency; so long as he holds onto his purely extensional account of what was done.

3. The Argument from Attribution

To describe an event as an action is to make an attribution to the agent. But just what the attribution consists in has separated action theorists. And not surprisingly, since the nature of action attribution will depend on the natures of action and agency: they define what we attribute and to what it is attributed.

According to Davidson's expression of agency, what we attribute to an agent when we attribute an action to him is an event under all true descriptions of it. Actions, for Davidson,
are extensionally conceived events. Which events we can attribute to an agent as his actions are those which satisfy Davidson's criterion. Only those events which stand in the right relation to the agent qualify as his actions: that is, only those events which are caused by an agent's intention count as actions of his. Therefore, while an action attribution attributes an extensional event to an agent as his action, the truth of the claim entails that the criterion was satisfied, which in turn entails that the event was caused by an agent's intention.

Davidson's criterion of action not only requires that actions be events with the right sort of cause, but also specifies something about the nature of the causing and so of the cause. This something is just what allows us to say that the event was intentional under at least one description. An event is intentional under a description if it was caused by an intention of an agent. This makes the causing an intentional causing - a causing with a purpose - that purpose being to satisfy the cause, the desire which motivated the action. That the cause is satisfiable implies that it has a representation of its object - that which would satisfy it. Intentions have both cognitive content and causal efficacy, and what they cause is what would satisfy them.

In specifying this about the nature of the cause of an action, an action attribution clearly specifies a limit to the nature of the cause. We attribute certain desires and beliefs -
those which cause and explain the action - with their cognitive content. Just as clearly, we do not attribute all possible desires and beliefs causally relevant to an extensional event. That is, we do not attribute an intention for every true description of the event. In attributing an action to an agent, we attribute a specific mental event, and so a specific cognitive content, and causal efficacy which causes that which will satisfy it.

Therefore, to take, as Davidson does, an extensional event as the thing attributed in action attribution without taking into consideration the nature and the limits of the agency, is, I will argue, both an over and an under attribution. The source of this difficulty for Davidson is the separation of the criterion of action from the expression of agency. It is just this separation which allows him the extensionality of the event, and which prevents him from accommodating theories of motivation and responsibility within his theory of action.

That the separation of the criterion from the expression is artificial will be apparent if we recall the function of the criterion. Its job was to get intention, and so purposefulness, and so agency into the picture. The criterion is what allows us to distinguish arm-raisings from arm-risings, purposeful behaviour from mere bodily movements. Purposeful behaviour is behaviour with a purposeful cause, and a causing which, in

\[\text{Davidson, 46-47, 120-121, 147, 195.}\]
standard cases, produces that which satisfies the cause. But what will satisfy a desire is not an event under all true descriptions. What will satisfy a desire is what figures in its representational content. That will be something like aspects of events - the extra-linguistic correlates of the descriptions which occur in the mental events which constitute intentions. To claim that actions are events under all true description is to attribute too much to an agent.

Davidson's extensionality is one kind of over-attribute. It is an over-attribute of descriptions of what one does as agent. Each new consequence of an action furnishes us with a new description of what was done. This is Davidson's accordion effect. Another way for an action ascription to over-attribute is for it to attribute too many actions. It is this charge which Davidson rightly levels at Feinberg. Feinberg's accordion stretches over events, and each new event causally related to an action presents us with a new deed. His sin is to multiply actions; his attributions, then, attribute too many actions to the agent.

Davidson's extensionality of action forces an under attribution to the agent by leaving the intentionality of the cause out of the action attribution. Remember, it is the expression and not the criterion which is attributed to an agent when we ascribe an action to him. Once the event which was caused by agency has been severed from the nature of its cause, and treated extensionally, there is no way to re-connect such a
thing with agency, and so with intention. To claim that we attribute extensional events to agents as their actions is an under attribution since it is not such a thing that figures, or could figure, in intentions. Davidson's extensionality forces an under specification of the intention entailed by an action attribution.

Just what the criterion of action was introduced for is frustrated by its separation from the expression of agency. The separation is what allows Davidson to get intention, and so intension, back out of the picture. But, with the separation, we face a dilemma: either we connect the criterion to the action so that it is part of what is attributed with action, in which case what is attributed to the agent is both the intentionality and the causality, and therefore not mere extensional causation; or, the criterion is not part of what we attribute with action, in which case action is supernumerary to all or most of the conceptual connections of action discourse. Davidson would find himself on the latter horn, but, as I have argued, theoretical considerations favour the former.

C. SOME IMPLICATIONS FOR A SEMANTICS OF ACTION

My arguments against Davidson's extensionality, if sound, have positive implications for a theory of action. I here sketch the alternative semantics of action suggested above, and which, I argue, meet the requirements on a theory of action. These brief remarks will be developed more fully as I develop a
theory of intending, intention and acting intentionally in the sequel.

Davidson, it would seem, has identified the wrong sort of thing as the action. An event is, according to the literature, supposed to be the sort of thing under no descriptive restrictions: extensionality supports non-equivalence of descriptions. I have argued that this semantics won't do for action. Unless we assume that the intensionality of intention implies a representational or intensional restriction on the action, it will be devoid of the intentional content without which theories of motivation, responsibility and attribution are unworkable. We need, therefore, some additional semantic room over and above a world of extensional events and ordinary event causality for the intentional content which we have found to be ineliminable from action discourse.

While Davidson is right in that to attribute an action to an agent entails attribution of causality, he is wrong, I have argued, to think that it is mere causal efficacy. That he does is evident in his separation of the criterion of action, and so of intention, from the semantics of action, or of what we do. With intention we who engage in action discourse accept some state of the agent which is both cognitive and causal. Therefore, we need a notion of cognitive efficacy or efficacious

---

42 Davidson, 105-203.
cognitivity. This, I claim, is just the role desire plays in our ordinary belief-desire psychology. Desires are just that explanatory concept, both cognitive and causal, which we need to explain action. And action is just that notion with which we attribute this cognitive efficacy to an agent, and therefore, along with it, the effect of that attribution, since attributions of action are attributions of agency.

Next, we need a notion which is the objective correlate of cognitive efficacy - that is, what such a cause causes. That will be whatever is the extra-linguistic correlate of the descriptions which occur in the intentions which cause action, and which provide the essential intentional content of actions. Descriptions correspond to properties or attributes. Therefore, the concept we need for action is not an event under a description, but an event-under-a-description, an essentially intensional item. Which description the action essentially embeds will be just that one which occurs in the causally successful intention.

Finally, the relation which holds between these essentially intensional items, must be intentional causation: cognitive efficacy or efficacious cognitivity functioning causally. This we accept when we accept that reasons are causes of actions.

This may be Dretske's notion of information.
We saw, in the exegesis of "Agency", that Davidson was trying to find an analysis of agency and action without having an account of intention. But, it seems clear that once we take intentions as the cause of actions, then the relation between an agent and his actions will be the relation between the agent's intention and his actions. The nature of the relation will be explicated by one of its terms - desire. Agency is that faculty of an agent by which his epistemic, teleological and intentional states produce behaviour. I turn therefore to the problems of intention, intending, and acting intentionally.
CHAPTER II
RATIONAL AGENCY

In this chapter I present a theory of rational agency in which will be found an account of intention, intending and coming to intend. Taking rational agency as basic, I diverge both from the tradition in philosophy of action which begins with Anscombe taking what an agent does as the basic case by which to understand intention,¹ and from Bratman's "methodological priority of future-directed intentions".² I take a middle road, or rather both roads, with agency - the forming of an intention which is efficacious of an action - as the centre of attention. I claim that we are less susceptible to distortions both of action and of intention if we maintain, what I argue is, their essential connections, and which are necessary if we are to maintain the normal inferences which attend both notions. I begin with the common ground for requirements on a theory of intention.


A. PROPERTIES OF INTENTION

There is a family of concepts which is connected with intention and which a theory of intention must accommodate. Among them we include motivation, rationality, commitment to action and the various modes of practical reasoning. In addition, there is the relation which intentions bear to actions. In this regard, there is some common ground. All who accept Davidson's thesis that reasons are causes and who accept common-sense belief-desire psychological explanation accept that desires and beliefs figure in the causal ancestry of actions. But how they figure, and their relation to intentions and intentional actions is a matter of dispute. With intention, we accept the concept of a state of agency which is causal of actions. What more we accept and need I will discuss by first considering what Donald Davidson and Michael Bratman have written on the subject.4

1. Davidson on Intention

In his paper "Intending" Davidson's concern is with the properties of intentions. Among these is the conceptual connection between intending to A and being committed to A-ing.


Commitment is a necessary ingredient of intentions: O intends to A only if O is committed to A-ing. It is this aspect, at least, of intentions which distinguishes them from mere desires, which lack this property. Since intentions contain commitment and desires do not, Davidson concludes that intentions cannot be reduced to desires. This marks a departure for Davidson from his earlier account of intention in "Actions, Reasons, and Causes",\(^5\) where intentional actions were seen as those which stand in an appropriate relation to the agent's desires and beliefs.\(^6\) But for the above reason, Davidson, in "Intending", has a non-reductive view of intentions. Intentions are like desires, in that both are of the genus pro-attitude, but intentions are all out or unconditional value judgments, whereas desires and wants are only prima facie value judgments.\(^7\)

The other property of intentions with which Davidson deals in "Intending" is what we may call purity. Purity is that property of intentions by virtue of which we may separate them conceptually from the other components of agency and of action. First, we can intend to A without A-ing. And, for Davidson, we can abstract intentions from any antecedent or consequent event

---

\(^5\)Davidson, 3-19.


\(^7\)See below, p. 69 f., where I argue that commitment is not an all or nothing concept.
such as deciding, deliberating, choosing, or any other form of coming to intend. This separation of intentions from the other components of agency, together with the claim that intentions are irreducible, is a view of intentions which we may call the *sui generis* view.

One way to appreciate the force of Davidson's irreducibility claim is to appreciate the following irregularity in the attempt to explain actions by some account of mental causation. According to the reductive view, an intention is analyzable as a desire-belief complex, and intentional actions are those which stand in the appropriate relation to these desires and beliefs. The difficulty with this view is that it does not, apparently, explain why it is that only some of an agent's desires are efficacious of actions. That is, it doesn't explain why we have mental causation when we do. Davidson is right to find this view of intention incomplete since it leaves out the notion of commitment which is constitutive of intentions. With commitment, intentions can do what mere desires can't: cause actions.

It is not clear that with the notion of commitment we have an explanation of agency that we didn't have with desires. In particular, we have as yet no explanation of why some commitments *don't* cause actions. Davidson's response could be that such cases may be explained by the intervention of some other component of agency, such as belief. But this move is
open to the reductive desire-belief theory of intention, and so Davidson's point does not tell against such a view.

Commitment and purity do pose problems for a reductive account of intending. I will show however that they are not insurmountable. But first, there is room for disagreement on Davidson's characterization of the property of purity. It is undeniable that one can intend to A without A-ing. This is only to accept the commonplace that one's mind can change with respect to some action, where this may be that one's degree of commitment can change: commitment is the sort of thing which can wax and wane. It would be a strength of a theory of intending that made explicit how this might work. However, intentions are not without essential, i.e., impure relations both to antecedents and consequents of the system whose intentions they are. As noted above, there is a conceptual entailment between intentions and commitment to action. That is, intentions entail restrictions on the efficacy of competing desires. To intend to A is to be in a state of agency such that competing desires, should there be any, are held in check, and such that one will, *ceteris paribus*, A. Therefore, intentions restrict which other intentions a rational system can have. As such, intentions are intimately, and inextricably connected to the rest of agency. One of the strongest connections is with rationality. What I suggest is that when we pick up all the relations between all the other concepts essential to agency, we will have intention.
If purity, for Davidson, means that intentions are not essentially related to the other components of agency, then, presumably, one could intend to A without having come to intend via some rational process between and among one's desires, beliefs and other intentions. If so, one could "parachute into" a state of intending to A; that is, just find oneself intending to A, with all that entails. But such a separation of intention from the rest of agency is incompatible with an important essential ingredient of intending. It must be rational for an agent to intend to A, where to be rational is for the intention to be compatible with the agent's desires, beliefs and other intentions. Therefore, what desires, beliefs and other intentions an agent has restrict which further intentions he can form. If so, intentions are, at the very least, not without inextricable connections to the other components of agency. Therefore, one could not just find oneself intending to A, where this means that agency has been circumvented. Intentions are not pure in this sense.

So far, then, we have found that a theory of intention must accommodate the following properties of intention: the unconditional value-judgement, rationality, and the possible separation of the intention to A, from A-ing.
2. Bratman on Intention

Michael Bratman deals, in his recent work on intention,\(^8\) with several of its aspects. Among them is the property of rationality by virtue of which intentions must be consistent with one another, a property which does not extend to desires. This demand on intentions allows what he has called the co-ordination of intentions in planning. Bratman also deals with the relation of what one intends to what one does intentionally. Though I cannot here do full justice to Bratman's work on these issues, I will sketch his position and raise some concerns I have about certain of his views which have relevance to this project. In the next chapter, when we will be better equipped to do so, I argue directly against one of his main points.

One of Bratman's strategies is to take planning as central to agency, and so to the concepts of action and intention. We are, he says, planning creatures. The co-ordinateability of intentions into plans is a result of the rationality constraint on intentions. Since intentions must be consistent, and desires not, intentions are, he says, irreducible to mere desire-belief complexes. Planning is a function of agency which cannot, Bratman argues, in consort with Davidson, be analyzed in terms of beliefs and desires.

The other main issue with which Bratman deals is the content of intentions. Bratman is concerned to preserve the

\(^8\)Bratman, op. cit..
conceptual connections between actions, responsibility and intentions. In particular, he is concerned with the implications of ascriptions of responsibility for the content of intentions and the relation between what one intends and what one does intentionally. According to what he calls the "tight fit" theory, there is an intimate connection between the content of one's intention and what one does intentionally, so that one does intentionally only what one intends to do. But Bratman finds that this threatens the connections we employ between action and responsibility, since one can be held responsible for what one does not do intentionally. Through a series of counter examples, the connections between what one intends and what one does intentionally are so weakened that one can do intentionally what one does not intend to do. Thus weakened, ascriptions of responsibility are preserved for cases where we want to hold someone responsible for what they did not intend, but nevertheless caused. To fill the gap left in the relation between intention and intentional action, Bratman introduces the notion of the motivational potential of an intention:

A is in the motivational potential of my intention to B, given my desires and beliefs, just in case it is possible for me intentionally to A in the course of executing my intention to B.  

Motivational potential is what connects agents to their actions.

---

9See below Chapter III, sec. A.

10Bratman, "Two Faces", 395.
We notice that Bratman's weakening of the connection between what one intends and what one does intentionally is like Davidson's separation of intention from the expression of agency. I resist this weakening for the same reasons I resisted Davidson's separation, and will argue against it after I have set out the theory of rational agency. For now, let us notice that once the separation is made, we find ourselves capable of attributing to agents full responsibility for events for which agency is responsible only in a diminished way. And, it is possible to recognize the diminution, and important for the theory of intention and action to mark that difference. The issue is not whether we should call such cases actions. Rather, it is that we find ourselves equipped to make important and subtle distinctions based on the relationship between the mind of the agent and those events which he effects in the world, and we should do so. After all, this is just what we must be concerned with as wielders of action discourse, and therefore just what an action theory should find.\textsuperscript{11}

The other criticism I have of Bratman, is a criticism of strategy. Planning cannot be central to action and intention when we can both act and intend to act without planning. This is not to deny the importance of the uses of reason, consistency and planning for mature agents. But, the strategy distorts the outcome. The view of intentions which I offer is a reductive

\textsuperscript{11} See below, sec. C.
one, but is nevertheless, I argue, one which satisfies all the acceptable requirements on a theory of intention.

B. A THEORY OF RATIONAL AGENCY

I shall develop a theory of what it is for an agent to intend - to have an intention - and to act intentionally. I begin with a simple case of action, and ask what particulars, properties, and relations are necessary for the agent to have so acted. To this end, it is helpful to construe the task as that of describing and constructing a system, both rational and teleological, which is capable of action. Such a system we may call an agent. In order to facilitate the task I introduce an action schema, which will be developed as the description of the system develops.

Consider the man, O, who releases the secret papers in order to embarrass the government. It is true of O, that he desires, \( d_1 \), to embarrass the government and believes, \( b_1 \), that the release is a means to that end. That he so acted means, among other things, that \( d_1 \) and \( b_1 \) caused, \( \rightarrow \), the desire, \( d_2 \), to release the secret papers. This is the causal compliment of Aristotle's practical syllogism. In terms of an action schema we have:

\[
\text{Fig. 1. } d_1 \& b_1 \rightarrow d_2
\]

This represents the mental side of things for O. Add to this a representation, by O, of the world of events which he effected
through his agency: the event of the government's embarrassment, \( E_1 \), caused by the event of the release of the secret papers, \( E_2 \), which was caused by \( E_B \), the event of O's hand moving. We now have:

\[ d_1 & b_1 \rightarrow d_2 \ldots \quad E_B \rightarrow E_2 \rightarrow E_1. \]

Thus we have, on the left hand side of the schema, a representation of O's mentality, and on the right hand side, a representation of the world. The ellipsis in Fig. 2 will be filled in as we proceed. It is important to note that as yet we do not have an action represented since the \( E \)'s, as they occur here, represent extensional events in the causal nexus. Before it can be called a schema of O's action, the right side will have to be articulated to bring in the essential intentional content of actions.

What we have so far represented of O's mental life are the propositional attitudes of belief and desire. Therefore, the \( d \)'s and \( b \)'s of the schema represent states of the agent with propositional content, different subscripts to represent different content. With desire we have a rich notion of a state of agency which is cognitive and motivational, and meant to be sufficient to initiate the processes of agency. We can ask what properties such a state must have in order to function as it does. First, it must have causal power to move the system
through the processes of agency,\textsuperscript{12} and to effect basic bodily movements by which agents affect the world in order to cause the states of affairs described in the propositional content of their desires. That desires are causal with respect to their content implies that the causal property of a desire is voided by the cognition that its object has occurred. Desires are self-extinguishing in this way. With the notion of an item which is causal of a state of affairs which is its object and which will void its causal power, we have both the notions of satisfaction of a desire by the recognition of the occurrence of its object, and a state of pre-satisfaction integral to the "pro-ness" of the pro-attitude of the desire. With pro-attitude we have the notion of being attracted toward, or cognizant of the attractiveness of, the object of the desire. This is the basis of value. And finally, with desire, we have a state of agency which the agent may report upon. So we have, with desire, these properties:

1. cognitive efficacy
2. voidability of causal power
3. satisfiability
4. pre-satisfaction
5. "pro-ness", or value, or attractiveness
6. reportability

With belief we have the concept of a cognitive state with content, and the capacity to affect desires, and through this, to effect other desires and states of affairs in the world. Beliefs can channel the cognitive efficacy of a desire by

\textsuperscript{12}This will be elaborated upon shortly.
providing the means to its satisfaction. O only desired to release the secret papers, and only did so, because he believed that would cause the embarrassment of the government. Although the function of beliefs is wide-ranging within agency, the influence of beliefs on states of affairs beyond the body is always transmitted through desire. That is to say that desire, being both cognitive and causal, is our representation of the idea of the points of interface between the cognitive and the physical at both ends of agency. This is our idea of agency. With belief, then, we so far have the following:

1. capacity to affect desires
2. capacity, in conjunction with desires, to effect other desires, other states of agency, and through them, states of affairs.¹³

We can now ask whether we have on hand a set of properties and relations of O's state of agency which is sufficient to cause his action. For it to be sufficient, d¹, b¹, and d² would have to be sufficient to constitute motivation. But it seems that as yet they are not. It is widely held that desires, or desire and means-end belief pairs constitute motivation. To identify a desire as a state with cognitive efficacy is to be committed to the view that desires are the source of motivation - the motive force. And, for a rational system with agency, a means-end belief is necessary to serve the desire in rational action. But, from the occurrence of a prime motivant, a d¹, or of a d¹ together with a means-end belief, a b¹, nothing follows

¹³This list will be added to shortly.
in the way of action without a belief to the effect that the
desired state of affairs is not or may not yet be the case.
Without this belief, one would normally, i.e., rationally, not
yet be in a state of motivation. Furthermore, one will be in a
state of cognitive efficacy toward an object state of affairs
only if one believes that the state of affairs is a possible
one. This is not to say that it is necessary to know how to
produce the goal, or to believe you have the means, in order to
be motivated, but only that the agent believe that the goal is
not impossible. This requirement constitutes part of the reason
why one cannot intend what one believes to be impossible.
Therefore, cognition of the way things are and of how they could
be is presupposed by motivation: agency, and so purposefulness,
presupposes cognition.

As an aside, consider briefly the state which is
categorized as having some of the properties of desire, but
because of the lack of the belief that the goal is a possible
one, is not yet a state of desirous efficacy. This would be a
pro-attitude, and so have the properties of pre-satisfaction and
"pro-ness" or value, but would lack causal power. We may call
such a state a wish or a hope, a state of the genus pro-
attitude, but lacking the cognitive efficacy of a desire. A
wish or a hope could become a desire, but in so doing would
cease to be a mere wish or hope. To corrupt a phrase, a wish
may be father to a desire, but only if efficacy is the mother.
And, a desire may degenerate into a wish or hope, indeed will
for a rational system, when the system acquires the belief that its object is not possible.

This gives us another function of belief. Beliefs have the power to void the efficacy of a desire. The belief that the desired state of affairs is impossible, or is current, is sufficient to cancel the causal power of the desire.

With this analysis of the principle agency state, we can see what more we need to add to the action schema. With the pre-supposed non-actuality and possibility beliefs, we must add a set of background beliefs. In addition, the fact that 0 acted on \( d_1 \) entails, for a rational system, the belief that so acting would not frustrate more important or more powerful desires of 0. That is, before the efficacy of a desire may go through, the question of consequences for the agent's other desires of satisfying the desire in question must be asked, and if not, is a departure from rationality, and one form of irrational behaviour. This entails that the mature agent (a system with more than one desire) has a set of desires which are ranked in two ways: according to causal power, and attractiveness. It also entails a set of beliefs about the set of desires and the rankings. We can add to the schema the set of desires \( <d> \), and a set of beliefs which includes the pre-supposed beliefs and the beliefs about desires, \( <b> \).

We are now in a position to describe the process of coming to intend. The first step of the process will be that point at
which \( d_1 \) becomes "active" for the agent. We begin, therefore with the set of desires, \( \langle d \rangle \), and the set of beliefs \( \langle b \rangle \). Desire \( d_1 \) will emerge from the set of desires either because of some cognitive cause, such as the belief that an opportunity has arisen to satisfy it - i.e., that the world has "lined up" in such a way that satisfaction is possible - or because of \( d_1 \)'s circumstantial or relative causal strength:

![Fig. 3](image)

This represents step 1, where \( d_1 \) "drops down" to engage in the subsequent stages of the action process. Step 2 of the process is for the occurrence of the activation of \( d_1 \) to prompt the system to evaluate the consequences for \( O \)'s other desires of acting so as to satisfy \( d_1 \). To repeat, this is a rationality requirement on the system:

![Fig. 4](image)

An answer to this question entails a ranking of \( O \)'s desires. If acting so as to satisfy \( d_1 \) has more value for \( O \) than satisfying any other of his desires, or more value than those desires which may be frustrated by so acting, then the answer to the question
about compatibility with the members of \(<d>\) will be positive, "+", and \(d_1\) will, under the impetus of its causal efficacy and because of this first validation, proceed, step 3, to the next stage. If the answer were negative, then the process of coming to intend will cease, if rational. This is to say that O's beliefs about the consequences of \(d_1\) proceeding in the process constitute an on/off switch for \(d_1\). The causal power of \(d_1\) may not be voided, but if not, it will be held in check by O's beliefs. This is to say, I will argue, that a negative answer at this point will prevent the system from intending \(d_1\). The negative answer, in the form of a belief is a gate preventing \(d_1\) from entering the further pathways of the system.

Since an action schema is also a rationality schema, where \(d_1\) proceeds despite an "off", we have a form of irrationality called compulsion. In these terms, compulsion is for a desire to be efficacious because it is too strong for the system. Similarly, we could have a system whose beliefs are too weak to prevent desires from being efficacious of actions. Or, an immature system may not have sufficient desires and beliefs to constitute competition for some desire, which therefore passes this stage unimpeded.

Returning to the process, we have \(d_1\) which has survived this first stage in the vetting process:
The occurrence of \( d_1 \)'s arrival here, on what I shall call the "I-line", prompts the rational system to ask, step 4, for implementation of \( d_1 \). That is, \( O \)'s beliefs concerning the means available to him, \( \langle b_m \rangle \), are queried:

From among this set, a set of beliefs about the means to satisfying \( d_1 \) is identified because they are believed by \( O \) to be the best available, because most effective. The members of this set are themselves ranked by effectiveness. This entails a set of second-order beliefs for \( O \). From among the set of most effective means, \( O \) selects, step 5, the means which he believes, \( b_1 \), to be most effective. Given that he desires the end, \( O \), at step 6, desires, \( d_2 \), the means. Next, he refers this desire for
means to his beliefs about his desires, step 7, and asks again, as in step 2, how the correlative event would affect his other desires:

This process requires that for each member of the set of means available, O's beliefs about his desires are queried as to the acceptability of the occurrence of the state of affairs which it represents, together with the occurrence of the state of affairs represented in $d_1$. This will involve a ranking of effectiveness and compatibility or acceptability of the events envisaged, by summing the values which each component has.

If the answer to the query at step 7 is positive for some means, then O, again if rational, will be caused, step 8, to desire, on the whole, that means. For our agent O, the means available to him is the release of the secret papers. Therefore, O is caused to desire the release, $d_2$, because it is the means to satisfying $d_1$:  

![Diagram](image_url)
If the answer at step 7 is negative, either because there are no means available, or because none of those available are acceptable to 0, then the causal efficacy of $d_1$ is voided, in the first case, or held in check, in the second, by 0's vetting beliefs.

In coming to desire, $d_2$, the release of the secret papers, 0 need not desire that event under all of its true descriptions. What is required of 0, if he is to be rational, is that he desire the means to satisfying $d_1$. Therefore, $d_2$ may occur for 0 under that minimal description - desiring the means to embarrassing the government, or desiring the means-ness of the release of the secret papers. Of course this event may be desired for other of its properties, under other of its descriptions. But, 0 may desire the means-ness of the event while desiring that it would not have other of its properties. 0 may believe that the release of the papers will be the same event as the cause of the loss of his job, or that which will stigmatize his family. But, the desire to embarrass the government may be so strong, for 0, that he desires the means
despite these undesired consequences. What he cannot do, if rational, is desire the means when that event under that description is more undesirable than the goal is desirable. What he must do, if rational, and so long as the vetting process has yielded positive results, is desire the means. So $d_2$ must have at least this minimal content.

Just as the occurrence of $d_1$ on the I-line prompted the system to ask "How?", so the occurrence of any desire for means may occasion the question of implementation. Indeed, it will occasion this question for any desire which the system as yet does not know how to implement, and where the system is rational. But for any system which is capable of affecting the world, there must come a point at which the question of implementation does not arise. That point, for our man O, will be the one where he "knows", either by learned patterns of behaviour or background abilities, how, for creatures such as we, to engage the world with his body. It is the point in the process of coming to act where further means do not enter.

O at step 9 asks, of his beliefs about means, how to release the secret papers, and repeats steps 5 to 7 at steps 10 to 12. It turns out that he can, $b_2$, move his hand, which holds the secret papers, toward the newspaper reporter, and so is caused to desire that, $d_3$, at step 13:
In order to satisfy \( d_3 \), \( O \) need not ask for means. It is sufficient for \( d_3 \) to be efficacious of some bodily event that it be vetted as above in the process of agency. That this is so, allows us to call \( d_3 \) a basic desire, or \( d_B \). Basic desires are those which effect the world directly through our bodies via the background of abilities we have to move our bodies. This must be true for some desires or agency could not emerge. Just as there are primitive actions antecedent to which we perform no other, there are mental events after which occur no relevant others before we act. This is represented in the schema as:

Fig. 10. \( \ldots \quad d_B \longrightarrow E_B \)

where \( E_B \) is the basic event in the world of, in our example, \( O \)'s hand moving.

Where an agent was successful through action, the relevant desires, screened and channelled by beliefs in the vetting process, will have caused those states of affairs in the world which satisfy the efficacious desires. That is, the efficacious mental events, with their representations of possible events,
will have caused a corresponding set of actual events in the world. Taking the bottom I-line of the left side of the schema, we may represent this as:

Fig. 11. \( d_1 \ldots d_2 \ldots d_B \rightarrow E_b \rightarrow E_2 \rightarrow E_1 \)

where, in our example, \( E_2 \) is the event of the release of the secret papers, and \( E_1 \) is the event of the government being embarrassed.

We do not, as yet (fig.'s 9-11), have represented an action schema. What we have yet to add is the notion of satisfaction. Satisfaction, I have argued, is a cognitive and intensional matter. What satisfies a desire is not some extensional event but some state of affairs or set of properties which is represented in the cognitive content of the desire. It is an event-under-a-description, an essentially intensional item which satisfies a desire. Therefore, before we can represent the matching of the causally efficacious desires with their corresponding objects in the world, we must construe the E's of our schema as the intensional items described above. Furthermore, what "satisfies" the instrumental beliefs of the schema will be the relation which holds, in successful actions, between these E's. We represent the satisfaction relation with '\(-S-\)'. A standard action will have the following form:
Fig. 12. A MODEL OF RATIONAL AGENCY: RAAG

The Mental

The World of
External
Events

\(<d>\): the set of RAAG's desires and intentions
\(d_1\): the desire which becomes causally active
\(<b>\): the set of RAAG's beliefs about \(<d>\)
I: the line of intending
\(<b_m>\): the set of RAAG's beliefs about \(<d>\)
b_1: the belief that (such and so) is the means to \(d_1\)
d_2: the desire for the means to \(d_1\)
b_2: the belief that (such and so) is the means to \(d_2\)
d_b: the basic desire
E_b: the basic event under the description in \(d_b\)
E_2: the event under the description in \(d_2\)
E_1: the event under the description in \(d_1\)

\(<-S>: the relation of intentional causation
\(\langle-S:\rangle: the relation of satisfaction

1: \(d_1\) becomes causally active
2: \(d_1\) is vetted by RAAG's beliefs about his other desires
3: \(d_1\) becomes an intention
4: RAAG's beliefs are searched for means to satisfy \(d_1\)
5: a belief, \(b_1\), about means emerges
6: RAAG is caused to desire the means
7: the means is evaluated by RAAG's beliefs about his desires
8: \(d_2\) becomes an intention
9: RAAG's beliefs are searched for means to satisfy \(d_2\)
10: a belief, \(b_2\), about means emerges
11: RAAG is caused to desire the means
12: the means is evaluated by RAAG's beliefs about his desires
13: \(d_b\) becomes an intention
14: RAAG's intention is efficacious
Thus we have a model of rational agency. And with the aid of the schema we can see how to separate these concepts. Rationality is what comes after $d_1$'s first appearance in the schema. The rationality of agency is that which serves the satisfaction or its motivating desires. Agency is $d_1$, plus rationality, plus $E_B$. It is efficacy of an event under a description according to the process described above. And Autonomy is non-interference with, and non-aberration of, the process and its constituents.

Of course the form of RAAG in 12 is not meant to be rigid. It can be as simple as an occurrence of a $d_B$ causing an $E_B$, or complicated by iterations of the module that is steps 4 - 8, the "means-module". This would be the case for complex rational agency.

In the full-blown schema 12, the bottom line of desires, $d_1$, $d_2$, $d_B$, is the line which represents that the causal power, as well as the other relevant properties of desire, are intact, by virtue of the vetting process of the agent's beliefs and desires. Therefore, intending begins only after the desire, $d_1$, has had its first vetting at step 2. That is, intending begins at step 3, where $d_1$ reaches the I-line. This marks the fact for the agent that satisfaction of $d_1$ is, according to his beliefs, on the whole desirable. Before the process of rationality begins, the agent merely desires but does not intend. Intending is the result of this first stage in the rationalization of desires. The test for whether a state of agency is an intention
is whether the agent can be in that state and not intend. I claim that at step 3 above, he cannot. The agent also intends at those points at which \(d_2\) and \(d_3\) reach the I-line - that is, wherever the vetting process yields a positive result. Thus we may say that intending - having an intention - is having a desire functioning causally and positively via the beliefs in the vetting process.

To report on the fact that one intends to A is to report that one's desire to A is in the vetting process and has had only positive vetting results so far. At any point in the process, of course, a query as to the acceptability of the events under consideration may yield a negative answer, in which case the causal efficacy of the desire, and so of the intention, is either voided or held in check, in a rational system. Therefore, one may indeed intend to A without A-ing. The schema thus elucidates how intending to A may not be sufficient for A-ing. It further elucidates the intimate relations between intentions and the rest of agency, the consistency requirements on intentions, the commitment which is constitutive of intentions, and the rationality of intentions. It also makes plain how plans and planning may function, since plans will be sets of intentions for future actions and, since they are intentions of a rational system, must be consistent. Future intentions may be "stacked" in \(<d>\) pending beliefs about opportunity and other aspects of planning for their satisfaction.
The I-line of schema 12 represents the intention with which, in our example, O acted. At $d_1$, O intended to embarrass the government, and so, by the nature of intentions, was committed to so acting. At $d_2$, O intended, not simply, or merely, to release the secret papers, but intended to release the secret papers so as to embarrass the government. The content of the I-line represents the content of the agent's intention. And so we see that the content of O's intention accretes as he learns how to satisfy his desire and then intends to so satisfy it. At $d_5$, O's intention is not merely to move his hand. It is to move his hand in just the way which he believes will be to release the secret papers so as to embarrass the government. This is the intention with which O acts.\(^\text{14}\)

Commitment is a concept which admits of degree. Just as the content of O's intention accretes as we move from left to right along the I-line, so his commitment may evolve accordingly. If, as O comes to intend to act, he discovers that the event he will cause will have further desireable properties, or fewer undesirable properties than he may have suspected, his commitment may increase. Conversely, his commitment may decrease as he discovers undesirable consequences and side-effects of the envisaged action. However, commitment is not just a function of desire. It is rather a function of the processional vetting of a desire. The further along one is, in the process, the fewer impediments stand in the way of the

\(^{14}\)This accretion of descriptions is the mental side of the accordion effect.
vetted desire, and efficacy is still "on" in light of the vetting beliefs. Therefore, commitment can increase, even though desire decreases.\textsuperscript{15}

With the full-blown schema, we have represented, in schematic form, all the various modes of practical reasoning, which will occur at the various nodes of the schema. Deliberating about how to act so as to satisfy a desire is the means-module, steps 4 - 8. Deliberating about whether or not to act, where this amounts to an evaluation of competing desires, is step 2, 7, etc.. Choosing to act may be where no clear ranking of desires emerges, and some prior motivant, such as the desire to act in some way, or this, together with a belief that opportunity will be lost if one does not act soon, causes a d to "drop down" to engage in the process. Choosing a means suggests that no clear ranking emerges, and, as above, the motivants already in the process cause the system to desire one of the equally acceptable means available. Willing to act is an interesting case. I identify will power with the causal power of desires. Therefore, one always acts out of will power. Yet the claim that one acted out of will power suggests that the motivants for the action itself were insufficient to move through the system, and some antecedent desire was involved in coming to act. Doing one's duty may be such a case, where the

\textsuperscript{15}See p. 44 f. above, for Davidson's account of commitment.
desire to do one's duty is the motivant, and not the desire for the action itself.\textsuperscript{16}

This, then, is the theory of rational agency, a rational process of vetting of a desire with cognitive efficacy. To assert truly that one intends to A is to assert that the mental process sketched above, which involves desires, beliefs and their relations, has taken, or is taking place. The phrase "the agent is intending to A" is true if and only if the result of the vetting process for the desire to A has been "go", and not "stop".

C. INTENTIONAL ACTIONS

We can see that the concepts of intending and coming to intend are concepts which qualify mental events. I turn now to an analysis of intentional actions - where intention is used to qualify actions - and the relation which holds between intending and intentional action.

First, there is the distinction between standard and non-standard actions. I claim that a standard action is one that is standardly caused by an agent, where this entails that the agent's intention was causally efficacious, in the way described above, and successful. That is, where an agent intends that such and so be the case, and where this intention causes that such and so is the case, then the agent has acted standardly.

\textsuperscript{16}I believe we have the means here to sort out the problems of weakness of will. I leave that task to a future paper.
Standard actions are actions which are caused, *in the right way*, by an agent's intention. Intentional actions are standard actions, and acting intentionally is for one's intention to be causal and successful. This is to say that there is no difference between acting, acting intentionally and acting standardly.

We may put this in terms of the action schema. For some event to be an action of an agent is for that event to be caused by an agent's intention where all the *matchings* are in place: the content of the efficacious desires are matched by the events-under-a-description, and the instrumental beliefs are matched by the causal relations between these. *This entails that the instrumental beliefs were true.* Therefore, an action is an event-under-a-description which is caused *in the right way* by an agent's desires and true beliefs. This is the analysis of action. What is ascribed to an agent in an action attribution is, first, that which satisfies his efficacious desires and beliefs - the intensional item represented in the agent's efficacious cognition - and which the agent caused. Secondly, we attribute to the agent the mental causation of the event-under-a-description, which entails that the schema is in order for the agent. All non-standard actions will be those for which the process I have schematized has failed in identifiable ways. We will find that we have, or could have, concepts for every type of aberration of the process of action.
In order to give an account of intentional action, we have found it necessary to appeal, at every point, to the mens of the agent. This is, I claim, because action is a diagnostic concept: action is a concept which entails a diagnosis of an agent - namely, his mental causation.

I have argued that actions are ineluctably intentional, and therefore intensional. This is to say that there can be no reference to an action that is not a reference to the ineliminable intensionality of the content of the agent's mental causation. Therefore, attributions of intentionality are not intensional, contra Davidson. Since action is tied essentially to the description(s) under which it was intended, any reference to the action is a reference to those properties which the agent desired. Thus attributions of intention and action are not intensional. Describe the action any way you like, if what you are describing is the action, then it is what the agent desired to bring about, and believed..., etc. Action essentially is, therefore, what the agent intended. Because actions are essentially intensional, substitution of non-equivalent descriptions of the action, in action attributions, is truth-preserving.

This may explain how it was that Davidson came to identify the wrong sort of thing as an action. An analysis of action discourse reveals that there is extensionality there. Davidson's conclusion is that actions are extensional. His
strategy is the same in "Causal Relations",¹⁷ where the presence of extensionality in causal discourse is enough to justify the claim that causation is an extensional relation between extensional events. But, that there is extensionality in action and causal discourse does not settle the matter as to the nature of the items involved.

I return at last to our agent 0 and the identity of his action. As I argued above, the intention with which 0 acted was the conjunction of the content of 0's efficacious desires. That is, 0 intended to embarrass the government by releasing the secret papers by moving his hand in just the way needed in his circumstances. Or, 0 intended to move his hand in just the way needed so as to release the secret papers so as to embarrass the government. Now we can ask, with Davidson, whether the basic bodily event was the action. Well, under the description "moving the hand" it is not. 0 does not intend to move his hand simpliciter, i.e., not under that limited description. What he intends is a bodily event with the right consequences. Nor does he intend the release of the secret papers, where this is treated extensionally. He intends to move his hand so as to release the secret papers so as to.... This brings out just how misleading Davidson's claim is that there are only primitive actions. Clearly, for creatures such as we, who can engage the world only with our bodies, actions will, of necessity, require a basic bodily event. But this is not the action. Where the

¹⁷Davidson, 149-162.
agent merely intends to move his body, the action will be the bodily movement under the description in his intention. For 0, because the event of his hand moving caused the event of the paper's release which caused the event of the government's embarrassment, we can describe his action as embarrassing the government. We can do this because the corresponding event occurred in the right way, which is to say, among other things, that the embarrassing was in the content of 0's successful intention.

Now suppose that the event of the government's embarrassment caused the event of the government's changing policy. What makes it false to attribute the change in policy to 0 as his action is that nowhere in his intention, let us suppose, does a desire occur for that event. We cannot attribute to 0's agency the change in the government's policy although we can attribute it to him causally. While intentions give us, as agents, a "way into" the world of events, just how far into that world our agency takes us, just how far the concept of action goes down the causal chain, is determined by the content of our successful intentions. Agency travels down the causal chain of events just as far as intention does. Of course, responsibility, sometimes being a matter of negligence, may travel further, but this does not entail that agency does. Therefore, contra Davidson, each consequence of an action does not present us with a deed.¹⁸

¹⁸See above, p. 18.
With this account of rational agency and therefore of action we are equipped to deal with some of the traditional problems of action theory. I turn, then, to applications of the theory to non-standard actions, the active-passive distinction, trying, and to a challenge to this account from the literature.
A. AGAINST THE SIMPLE VIEW

As was noted above, Bratman has a view of intentional action which separates the content of the intention with which an action is done from what is done intentionally so that one can A intentionally without having intended to A so long as one intended to B for some appropriate B. This is a weakening of the connection between intention and intentional action and so runs contrary to RAAG. Bratman defends this separatist view of intentional action by an argument against what he calls the "Simple-View", which states that:

for me intentionally to A I must intend to A; my mental states at the time of action must be such that A is among those things I intend.¹

I have been arguing against this separation of intention and intentional action and so have been defending the Simple View. I will give Bratman's argument against the Simple View and show why I find it unconvincing.

The argument consists of three examples the third of which is meant to do the trick. In case 1 we are asked to imagine someone, let's call him M, playing a difficult video game which requires firing a "missile" at a target. M is quite skilled at it but nevertheless is doubtful of success. He aims and fires. As it happens, he "succeeds", says Bratman, in just the way he was trying, which means:

1. hitting the target was what he wanted to do
2. the hitting depended on his skills and so was not a matter of luck
3. his perception of the hitting terminated his attempt (and so it was not inadvertence)

It is clear to Bratman that even though M is doubtful of success, and is trying, that if he "succeeds", he does so intentionally. If so, then on the Simple View he must have intended to hit the target which, for the Simple View, is a thoroughly acceptable result.

Now suppose that, in case 2, a second game of this type is added, and our ambidextrous M plays them simultaneously. Again because they are equally difficult, he is doubtful of success at either game, and equally doubtful since he is equally skilled with either hand. Suppose he misses target T2 but "succeeds" in hitting target T1, where conditions 1 - 3 above obtain. If so, says Bratman, he hits T1 intentionally. On the Simple View, M must have intended to hit T1. If so, then because the case is entirely symmetrical with respect to T2, he must have intended to hit T2 as well, even though that intention was not successful. So, on the Simple View, M intended to hit each target.
Finally, in case 3, we are asked to imagine that these two games have been linked so that it is impossible to hit both targets; that if both are about to be hit, the game shuts down and the player looses, and all of this is known to M. Still, there is a reward for hitting either target and since it is difficult to hit either, M decides to play both games simultaneously: he reasons that the risk of shutting down the machines is outweighed by the increased chance of hitting a target. So he tries to hit target 1, and tries to hit target 2. Suppose that M "hits" target 1, where conditions 1 - 3 above apply. Therefore, just as in the first two cases, he hits the target intentionally, says Bratman. On the Simple View, he must have intended to hit target 1. But given the symmetry of the case, M must have also intended to hit target 2. And here is the problem for the Simple View. Since M knew that he could not hit both targets, if he intended to hit both, which he must have according to the Simple View, he would be criticizably irrational, since his intentions would fail one of the rationality constraints on intentions. It is a requirement of rationality that intentions be strongly consistent:

My intentions are strongly consistent relative to my beliefs if all my intentions could be put together into an overall plan that is consistent with those beliefs.²

Since M fails this strong consistency requirement, if the Simple View is right, he is guilty of irrationality. But Bratman sees

²Ibid., 380.
no reason to so charge the player. The strategy he employed maximized his chances of winning given the difficulty of hitting either target. Therefore, the Simple View is false: M did not have both the intention to hit T1 and the intention to hit T2. But if so, then since the case is symmetrical with respect to both targets, M had neither intention, and so the Simple View is false:

The Simple View imposes too strong a link between intention and intentional action, a link that is insensitive to differences in the demands of practical reason.  

I think that some of Bratman's premises are true but that his conclusions are false. In arguing against Bratman I will accept the strong consistency requirement, since intentions are rational items, and I will agree that the game player in this third case was not guilty of a form of irrationality. All the same, I defend the Simple View, which in consequence I rename the Unified View.

The force of Bratman's argument against the Unified View turns on his claim that M hits T1 intentionally in case 3. If M doesn't do it intentionally, then of course there is no challenge to the Unified View. I shall argue that we have some good reason to deny intentionality to the hitting of T1 in case 3, and that the consequences of such a denial are not as serious (or as counterintuitive) as at first they may seem. According

3Ibid., 383.
To begin to cast case 3 in terms of the action schema it will be helpful, and I think unobjectionable, to say that M is no longer playing two games. He is playing and trying to win one game, the rules of which Bratman has provided. We assume that M desires, \( d_0 \), to win and believes that to win is to hit T1 or T2 but not both, and so desires, \( d_1 \), to hit T1 or T2 but not both. M comes to intend to do this at \( d_1 \), step 3. Since M is a rational system, the arrival of \( d_1 \) at step 3 - the intention-prompt - prompts M to search his beliefs about means for how to achieve such a hit, step 4. Now, M believes that in order to win he must hit either T1 or T2 but not both. But, since hitting either is difficult for M, he is doubtful of hitting either. Therefore, it is rational for him to believe that the best means available to him to satisfy \( d_1 \) is to try to hit both T1 and T2. Furthermore, he believes, that to try to hit T1 and T2 is to aim and fire the guns at the targets and so believes, \( b_1 \), that to do so is the means. At step 8, he comes to intend to so aim and fire. So far, there is no violation of the requirement of strong consistency on intentions, since he does not intend two incompatible actions, and does not have inconsistent intentions: he can intend to try to hit both targets without inconsistency since intending to try is not to intend to do. Next, M asks for means again, and believes, \( b_2 \), that to move his body in some certain way, in his circumstances, will cause the guns to be...
aimed and fired at the targets, and comes, at step 13, to intend, \( d_b \), to do so. The basic desire, \( d_b \), causes a chain of events under the following descriptions: M's body moves, \( E_b \), in just the way described in \( d_b \), which causes the guns to be aimed and fired, \( E_2 \), which causes a missile to hit T1, and another to miss T2. The schema of this is as follows:

Fig. 13. RAAG PLAYING THE VIDEO GAME

- \( d_1 \): the desire to hit T1 or T2 but not both
- \( b_1 \): the belief that to aim and fire the guns at the targets is a means to satisfying \( d_1 \)
- \( d_2 \): the desire to try to hit T1 and T2, i.e., to aim and fire at both
- \( b_2 \): the belief that to move his body in just the right way will aim and fire the guns at the targets
- \( d_b \): the desire to do so (as in \( b_2 \)) move his body
- \( E_b \): M's body moving in the way he desired \( d_b \)
- \( E_2 \): the guns being fired at both targets
- \( E_1 \): the event of T1 being hit and T2 not
Now, according to the Unified View M hits T1 intentionally only if he intends to hit T1, which causes the hitting. But M does not so intend. He intends, \(d_1\), to hit T1 or T2 but not both which is not to intend to hit T1. And, M intends to try to hit both T1 and T2, which again is not to intend to hit T1. Therefore, on the Unified View, M does not hit T1 intentionally. Nor does M hit T1 intentionally according to RAAG. What M does intentionally is what is caused in the right way by his efficacious intention. At \(d_8\), step 13, M intends to hit T1 or T2 and not both by trying to hit both T1 and T2 by moving his body in just the way he believes will do the trick. All of \(d_1\), \(d_2\), and \(d_8\) are matched by E's which satisfy them, and similarly for \(b_1\) and \(b_2\). Therefore, we can say of M that he intentionally hit T1 or T2 but not both.

To deny that M hit T1 intentionally is not to deny that the hitting was the result of M employing his considerable skills. Nor does it entail that M does not deserve credit for hitting T1. The claim of unintentionality is the claim that the action has deviated from the standard. The standard case of hitting T1 would be where the agent intends to hit it and believes that such and so is the means to hitting it, where these comprise an intention which is successfully efficacious of an event of the missile hitting T1. Case 3 is a deviation of the standard in two respects: first M has no such intention; second, M's instrumental beliefs are not functioning standardly, since he is doubtful of success. I turn to this issue in the next section.
It will be apparent that the method employed here is in sharp contrast to Bratman's strategy. The strategy of taking agency as basic, as the place to begin sorting out hard cases, is clarifying and so helps sort out conflicting intuitions.

B. TRYING

Given the above analysis of case 3 what should we say of M while he was en acte? Since he intentionally hit T1 or T2 but not both, we can say that he was hitting T1 or T2 but not both. We can also say of him that he was trying to hit T1 and trying to hit T2. And M would agree that at the time he was trying to hit T1 and not hitting T1, since he was doubtful of success. To report that one is trying is to report that efficacy is "going through" despite one's lack of confidence in one's means. Thus we have an analysis of a trying which admits several distinctions. First, there is a trying which may or may not be a doing. Where an instrumental belief is tentative and yet efficacy proceeds despite the lack of confidence, the agent comes, at step 6 or 11, to a "doubtful desire" for the means. This is a case of tentative cognition. Furthermore, where an instrumental belief is tentative, its influence on the process of coming to intend may yield tentative causation, where the agent engages the world haltingly. Secondly, there is a trying that is not a doing, where to try is to fail by having one's desires be efficacious under the impairment of a false instrumental belief. When someone intends to A but mistakenly
B's because of a false instrumental or identity belief, we can say of him neither that he was A-ing, nor that he was B-ing; what we can say of him is that he was trying to A but mistakenly B-ed. Similarly, when someone intends to A but accidentally B's because of a false belief, or lack of relevant true belief, we can say of him that he was trying to A. These cases of trying share a diminution either of cognition or of efficacy: cognition may be "on" but tentative; efficacy may be "on" but tentative; efficacy may be "on" but impaired. Therefore, we ought to be reluctant to say of someone who is intentionally A-ing that he is trying to A. He is not trying; he is doing.

Since M was doubtful of success in both cases 1 and 2, we can say of him that he was trying to hit T1. Assuming that M's trying is a case of tentative cognition, it is a deviation from the standard. Therefore, we have reason to deny that M hit T1 intentionally in cases 1 and 2. It is clear to Bratman's intuitions, but not so clear to mine (and I don't think this should count for much), that M does hit T1 intentionally in cases 1 and 2. But it is not enough to guarantee intentionality that the target was hit, since this fact does not guarantee that the hitting was due to agency. We cannot just assume success. Where it turns out that the agent's beliefs were true and that therefore efficacy was successful in just the way the agent envisaged, i.e., where Bratman's condition 2 obtains, we would attribute intentionality. Such a case would be a trying which

---

4See below, sec. C(1).
was a doing. But where all we know is that the agent was trying to achieve a result which was forthcoming, the question of intentionality is inconclusive. With the inconclusivity of trying we can attribute not unintentionality, but non-intentionality.

On this analysis, to report that you are trying is to report on the inconclusivity of your instrumental beliefs. To attribute a trying to another is to attribute either inconclusivity due to a "doubtful desire", or non-efficacy due to a false instrumental belief.

On the theory of rational agency which I am presenting, in order for an action to be intentional, it is sufficient and necessary that agency functioned in an appropriate way in the causal genesis of some correspondingly appropriate event, where the appropriate way is for the agent's desires and true beliefs to function causally in the process of vetting and adjudication which I have outlined. In standard cases of action, the process will include, among other things, instrumental beliefs functioning in their particular way. What Bratman has described in cases 1 and 2 is something less than this. In cases 1 and 2 the player does not believe that he will be successful since he doubts his means. But, since his desire to win is strong enough, he chooses the best means available which is to employ his considerable skills: he intends to try. In other words, the instrumental beliefs which would otherwise appear in an action schema do not enjoy M's full confidence: therefore, the
diagnosis we must make of M's cognitive efficacy is that it has not functioned standardly. So, while M intends to hit T1, and while his instrumental beliefs turn out to be true, so that his efficacious desires are matched by states of affairs which satisfy them, and his beliefs are matched by the relations between these states of affairs, we may deny intentionality, so long as doubt remains, as a recognition of the inconclusivity which accompanies tryings. The importance of marking this alteration of agency will be reinforced when we come to discuss excuses and responsibility, where the diagnosis of the mind of the agent is of central importance.  

Support for this view of trying as a cognitive concept comes from the fact that in Bratman's case 3, M can intend to try to hit T1 and T2 without violating the consistency requirement on intentions. The intention to try to hit T1 is not incompatible with the intention to try to hit T2, given M's beliefs about the impossibility of hitting both T1 and T2, since to intend to try to do both is not to intend to do both, nor is it to desire to do both, nor to believe that you will do both. Trying is the first of the ways I will examine in which the left-hand side of the schema - the agent's mens - can be out of order.

The above analysis of Bratman's case applies, mutatis mutandis, to Davidson's claim that one can do intentionally what

5This analysis of trying can be applied with some success to Analysis "Problem" No. 16. See Analysis 38, 113.
he doubts he is doing. In his example, a man, in writing heavily on his page, is intending to make 10 carbon copies. He does not know or believe with any confidence that he is succeeding. But, since this is what he wants to do, then, if he does "produce" 10 copies, he does so, say Davidson and Bratman, intentionally. On my analysis, he is trying to make 10 copies, and again, the question of intentionality is inconclusive. The result we may call good fortune, and his action a non-intentional making of 10 copies.

Part of the purpose to which Davidson puts the above example is to show that one can do intentionally what one does not believe one is doing. Therefore, there is no such belief requirement on acting intentionally. But if one can act intentionally without believing that one is, then, say Davidson and Bratman, one can intend to A without believing that one will A.7 This is to argue against what Bratman calls the strong belief requirement on intending to A. According to the weaker version of this thesis, if one intends to A, then one must believe that one will A. On the stronger version, to intend to A is to believe that one will. The thesis is meant to explain the oddness of remarks like the following: "I intend to go to the concert, but I may not go."8 Davidson and Bratman take

---

6Davidson, 50, 60, 91 ff.

7Ibid., 90 ff.; Bratman, 383 - 385.

examples like the above to defeat both versions of the strong belief requirement on intending and doing.

The view I have been arguing for may seem to include the strong belief requirement, but it does not. I have argued that in order to act intentionally, one's instrumental beliefs must be true and function for the agent in a standard way. This entails that the agent believe that he has the means to his goal. In a standard case of hitting the video game target, the shooter believes that if he moves his body in just the way he has in mind, the missile will hit the target. But this is not to believe that he will hit the target, although this latter belief is entailed by the former. Therefore, my requirement on intentionality is weaker than either version of the strong belief requirement.

The entailment relation is important for another puzzle of agency. If someone intending standardly to A were to reflect on his state of mind, he might come to believe that he will be successful. Therefore, if he were to predict his future behaviour, he would say that he will A. This explains the oddness of one saying "I intend to A but I may not", an oddness which led some theorists to the strong belief requirement on intending to A. Since intending implies confidence in success, and commitment to action, intending to A entails the belief that one will A. To say that you intend to A when you also believe that you may not be successful, either because you believe you cannot, or are not able, is to mislead the hearer.
C. NON-STANDARD ACTIONS AND EXCUSES

I have just given an account of one form of aberration in rational agency which we call a trying which is not a doing. It will be a further test of the theory of rational agency that it facilitates the legal and moral distinctions between standard and non-standard actions. Non-standard actions are those cases in which excuses, and so defenses, apply. Here I deal with only some of these distinctions.

1. Mistake

As I noted above, a mistaken action is a case of cognitive efficacy "going through" under the impairment of a false instrumental belief. Since it is an impairment, mistakes are unintentional. It will be helpful to show how this comes out in terms of the action schema. So we take the following example:\(^9\) Suppose that I desire \(d_1\), to meet Groucho Marx and come, at step 3, to intend to do so. I believe that Groucho is next door and so believe, \(b_1\), that going there is a means to meeting him, and therefore desire, \(d_2\), to go next door, and come, at step 8, to intend to do so. I believe, \(b_2\), that if I perform the familiar movements I will go next door and come, at step 13, to intend, \(d_3\), to so move. At \(d_3\), I intend to meet Groucho by going next door by moving in the old familiar ways. This "intention" is efficacious of the following events under these descriptions:


\(^10\)This is an example of Professor Coval's.
Eₜ, the event of my moving in the way I have in mind in dₜ which causes the event, E₂, of my arriving next door which causes the event, E₁, of my meeting my neighbour.

**Fig. 14. RAAG MAKES A MISTAKE**

The Mental

The World of External Events

I: d₁ → d₂ → d₃ → Eₜ → E₂ → E₁

**d₁:** the desire to meet Groucho
**b₁:** the belief that going next door is a means to meeting Groucho
**d₂:** the desire to go next door
**b₂:** the belief that walking next door is a means to going next door
**d₃:** the desire to move in that way necessary to walk next door
**Eₜ:** the event of my walking that way
**E₂:** the event of my arriving next door
**E₁:** the event of my meeting my neighbour

In this case b₁ is false. The intention with which I act is the intention to move in just that way in order to go next door in order to meet Groucho. But because b₁ is false, my goal is not forthcoming. That is, not only do I not meet Groucho, but I do not move in just that way in order to go next door in
order to meet Groucho. Therefore, neither \( d_1 \), \( d_2 \) nor \( d_B \) is satisfied, nor are \( b_1 \) and \( b_2 \) satisfied. Consequently, what I do in this case is mistakenly meet my neighbour. Since it was the false belief, \( b_1 \), that, together with \( d_1 \), caused me to desire, \( d_2 \), to go next door, I do not truly desire to go next door. It is a "false desire". If so, then \( d_2 \) at step 8 is not an intention. It is a misintention.

On Davidson's analysis of mistakes, making a mistake is doing something else intentionally.\(^{11}\) That is, there will be a description of what the agent does under which it is intentional. For RAAG this is false. Now it may seem that \( E_B \), the event of my moving in some way, is intentional. It would be if I desired to move in that way simpliciter. But I do not. The content of my intention at \( d_B \), step 13 is not just to move my body in some way. It is to move in some way so as to arrive next door so as to meet Groucho. But to move the way I do is not to move so as to come to meet Groucho. Therefore it is unintentional: in mistakes the basic bodily movement so described is unintentional.\(^{12}\)

The reason we recognize mistake as an excuse is obvious, since making a mistake is not for agency to have functioned standardly. If not, one ought not be held responsible for one's

\(^{11}\)Davidson, 45.

\(^{12}\)This analysis of mistake applies to Davidson's examples, which include spilling the coffee, the officer sinking the Bismark, and Hamlet killing Polonius. Davidson, 44 - 45.
mistakes to the degree to which we are in standard cases, since how agency functions in the non-standard case is entirely different than in the standard case. Where some standard of care applies which you do not meet and should have met in that you should have been more careful about your epistemology, about which beliefs you have, and, therefore, about which beliefs you allowed to function in consort with desires in action, then you are criticizable for the breach of the standard, and may be held responsible for the consequences. This is typically the case where the consequences of acting out of false beliefs are negative. To fail to meet a standard of care in the above way is one way of being negligent. It is to be negligent with respect to your beliefs.

2. Accident

With mistake, what cancels agency is the false instrumental belief. In accident, what does the cancelling is the intervention of another cause. As Davidson says "...if I spill the coffee because you jiggled my hand, I cannot be called the agent." The diminution of agency involved in accidents is either a lack of true instrumental beliefs, or the presence of false instrumental beliefs. In the above example, I may either not foresee that you will jiggle my hand, or I may falsely believe that, e.g., if I hold my cup away from you, the coffee won't get spilled. In either case I cannot be called the agent of the spilling, since I did not intend it: there is no desire

13Davidson, p. 45.
to spill the coffee, and no belief that it will spill. To illustrate, I will schematize the following example. Suppose I intend to shoot my donkey and believe that aiming and firing the gun at him will shoot him. But, at the moment of the shooting, my neighbour's donkey steps into the line of fire and tragically is shot. In this case I accidentally shoot my neighbour's donkey, since I lack both the belief that he will intercept the fire and the desire to shoot him. This differs from mistakenly shooting him, which would be to falsely believe, for example, that that donkey is mine, and not my neighbour’s. But as with mistake, in accident none of the efficacious desires and beliefs are satisfied.

\[14\] This is an example of J.L. Austin's.
Fig. 15. RAAG ACCIDENTALLY SHOOTS HIS NEIGHBOUR'S DONKEY

3. Compulsion

A compulsion is a desire whose efficacy is so powerful that it can override the rational processes of agency. That is, at the point in the rational process of coming to intend where desires are vetted by beliefs, a compulsive desire can proceed into the further channels of agency despite an "off" in the form of a belief that acting so as to satisfy the desire will be, on the whole, damaging to the system, where this means that more important desires of the system will be frustrated. A
compulsive action is what results from this sort of aberration in the system. Since agency has broken down in this way at that point in the process where desires are rationalized, compulsions do not acquire the status of intentions, and compulsive actions are not intentional actions. The schema for RAAG explicates the phenomenon of compulsion.

There are two related phenomena which can occur at the point in the process where compulsions violate RAAG. The first is the case in which a desire makes its way past the vetting process at step 2 not because it is too powerful to hold in check but because the other desires and/or beliefs of the system are too weak. The second case is one in which a desire becomes efficacious, not through rationality, but because the system lacks further desires with which the first competes. These may be two ways in which a system can be immature. Since they are both aberrations of RAAG, they may justify the denial of intentionality and responsibility.

4. Involuntariness

As I remarked in the previous chapter\textsuperscript{15} the theory of rational agency is also a theory of autonomous rational agency. Autonomy is for RAAG to function standardly unimpeded from without. In particular, it is for the first stage to occur, where \( d_1 \) "drops down" to engage the rest of agency, without interference, either from the environment in which the system

\textsuperscript{15} See above, p. 67.
finds itself, or from another agent. With such interference we have a case of involuntariness, of which we may distinguish two types. The defense of necessity applies to cases in which the environment is such that it forces a desire to become efficacious which would otherwise not have been. Given the set of desires and beliefs you have, the world may line up in such a way that you are forced to do something you would otherwise not have done, in order to preserve some greater good, for example your life or that of another. Throwing the cargo overboard in a storm in order to keep from sinking is a case of necessity, and would be excused as such.

The second type of involuntariness is for another agent to force you to do something that you would otherwise not have done. We may distinguish this from necessity by calling it a case of coercion. To be coerced is to be forced to have a desire become efficacious, given the rest of your desires and beliefs, which would otherwise not have been efficacious. For example, the gunman forces you to desire to give him your wallet, and you do desire this, but only because you believe that if you do not comply, you will be harmed. Given your desire to live, and your desire not to be harmed, you come to desire to hand him your wallet since you believe that that is a means to satisfying those more important of your desires.

It is important to notice that in contrast to compulsion, necessary actions and coerced actions are intentional actions, despite being involuntary. They count as intentional, for RAAG,
since the system functions as it should. You intend to throw the cargo overboard, and you intend to hand over your wallet. Given the situations in which you find yourself, these are the rational things to do. There is no diminution of agency due to false beliefs or compulsive desires, but only that your options are so severely limited, that you are forced to act in these particular ways. Your defense is that you would not have so acted had the probable consequences of not doing so not have been so dire.

D. THE ACTIVE - PASSIVE DISTINCTION

So far, RAAG has provided an account of what it is for agency to function standardly in the causal genesis of some state of affairs: we have standard actions. RAAG has also demonstrated some of the ways in which agency can function non-standardly in the causal origin of an event: we have tryings, mistaken actions, accidents, etc.. These cases have themselves been distinguished by the different ways in which agency has functioned non-standardly. There are several other ways for agency to be involved in the ancestry of an event which is not an action but for which an agent may be held responsible to some degree. These cases are sometimes identified by the active-passive distinction. In this section I will analyze several ways of being passive. As will become apparent, the term "passive" is misleading. Some of the cases I will describe which do not count as actions include several of the components of agency, and others will include agency, although not of the events in question. The purpose of the analysis is that with
RAAG we can distinguish several ways of being "passive". Since we can distinguish several ways in which agency can function in the ancestry of an event which is not an action, we have a way of assigning varying degrees of culpability for the corresponding event. We should expect that for every different mens we can diagnose, we could assign a different degree of responsibility, and do so with good reason. The case I will use is the familiar killing - letting die distinction.

According to RAAG, in order to count as a killing, it is both sufficient and necessary that the agent desire the death and believe that doing such and so will cause the event of the death, where the intention which is comprised of these desires and beliefs is efficacious, in the right way, of the event of the death. To begin to set up the contrast with letting die, we shall call killing the deliberate provision of causal sufficiency, in the circumstances, for the death.

It is necessary in order to count as a case of letting die that the agent be able to prevent the death and that he believe this or should have believed it. Suppose that an agent could have prevented a death but believed, on the best available evidence, that he could not, having met the required standards of care with respect to his beliefs. If it is not the case that he should have believed it possible to prevent the death, then we would not call this a case of letting die.
A case of letting die may be called an act of omission. To ascribe to an agent an act of omission is to describe his agency in terms of a relation that it should have had and didn't. That is, it is to say of the agent that he should have acted in some way which he did not and should have done so because of the appropriate desires and beliefs which he should have had.

We can distinguish two broad categories of letting die according to the presence and absence of the following desire and belief:

(1) the desire for the death
(2) the true belief that there is, or will be or could be a dying and so a death

To count as a case of intentional letting die, (1) is necessary. Without the desire for the death, all other cases of letting die are cases of negligent letting die.

Within these two categories, we can distinguish differences based on the four possible combinations of this desire - belief pair, (1) and (2).

1. Intentional Letting Die
   a. The straightforward case
      (1) desire for the death
      (2) belief that death is imminent

      This is the case we may call intentional letting die where
the agent desires the death and believes that the situation as it is is sufficient for the death. Since he desires the death, the agent does nothing to interfere with the causal sufficiency, either by standing fast with the purpose of not interfering or by doing something else. In such cases, the agent may be criticized for having bad desires, or for failing to intervene to prevent the death. To take an example\textsuperscript{16}, suppose that you are canoeing downstream toward the drowning man. You would let him die in this way if either of the following is true: you are paddling and you believe that one more stroke will bring you within grasp and so, since you desire his death, you do that with your body which will not interfere with the present causal sufficiency. Or the above is true, and you paddle in some way other than the way which would have let him save himself. This case of letting die is distinguished from the next by the belief here that non-interference is necessary in the circumstances for the death.

b. The hard case

(1) desire for the death
(2) belief that death is not quite imminent

In this case, the agent desires the death but believes that causal sufficiency does not yet exist for the death, and that agency is required in order to achieve the sufficiency. In our example, such a case would be where you believe that should you not paddle, or act in some other relevant way, the drowning man will come within reach of your canoe and so save himself. Since

\textsuperscript{16}This may be Dick Sikora's example.
you desire the death, you would let him die in this way if you, for example, take one more paddle stroke which you believe is necessary in the circumstance in order for there to be a causal sufficiency for the man's death. Thus, this type of letting die differs from the former in that in the former non-interference is necessary, while in the latter interference is necessary. In this case, the agent is criticizable for bad desires, and for acting so as to contribute to the cause of the event. To be guilty of this is to be guilty of a more serious offence than in the former case, since agency is employed here in a way necessary for the death.

2 Negligent Letting Die

a. Pure Indifference

(1) no desire for the death
(2) belief that death is imminent

The type of letting die which we may call pure indifference is characterized by the belief that, e.g., the death will occur, where such a belief does not "call up" or activate a desire to prevent the death by employing agency. That is, the belief does not function instrumentally for the agent. One guilty of pure indifference is criticizable, not for having bad desires, but for not having the right desires, given his beliefs, and therefore is criticizable for not acting so as to prevent the death.

b. (1) no desire for the death
(2) false belief that death is not imminent
It may seem that in order to count as a case of letting die, it is necessary that the agent believe that the death is imminent. But an agent can be guilty of a negligent letting die when he does not meet some standard of care with respect to his beliefs which is required of him. That is, he does not believe that death is imminent and should have so believed, perhaps because he has been charged with that responsibility. In our example, an inattentive lifeguard or river-keeper would be criticizable for the state of his cognition, for not having the right beliefs, the assumption being that, had he known of the drowning, he would or should have acted so as to prevent it.

c. (1) desire the death
    (2) false belief that death is not imminent

In this case of letting die, the agent is negligent with respect to the state of his cognition since he falsely believes that death is not imminent and should have known. In contrast to the previous case, the agent desires the death and so had he truly believed that death was imminent he would have approved and not acted so as to prevent the death. He is therefore criticizable for not having the right belief and for having the wrong desires. The presence of the desire for the death makes cases of type 3 prima facie more morally repugnant than those of type 1 and 2.

We may identify a variant of type-a negligence which counts as a defence against the charge of letting die. Where an agent does not desire the death and does believe that death is
imminent but does not act so as to prevent the death, then depending on his reasons, he may be excused. Where the cost to the agent would be so great that he could not be expected to intervene to save the drowning man, he may be excused from the charge of letting die. Where the agent fails to intervene out of fear, culpability may be reduced.

In all cases of letting die the role which agency plays in the ancestry of the death is less contributory to the death than is a case of killing. Therefore, letting die is less morally repugnant than is killing. Even in the problematic case, where the agent acts so as to prevent an event which would have prevented the death, the agent’s behaviour is less repugnant than a case of killing not because of any difference in desire between the cases, but because he employed agency in a very different way. To kill is to contribute the causal sufficiency for the death. To let die in the problematic case is to do what is necessary in the circumstances in order that causal sufficiency exists. The difference can be brought out if we imagine two people, the first of which goes about producing harmful events, and the other which goes about looking for imminent harmful events and either ensures that they are not interfered with or ensures that they occur by removing an impediment.

The central notion of RAAG, which is that action is a diagnostic concept, has allowed us to distinguish several types of letting die with their accompanying differences in moral
repugnancy. In all of the above non-standard cases, we have been able to make and maintain fine-grained distinctions in the mental causation involved. Since actions are cases of efficacious mental causation, with these fine-grained distinctions we have a foundation for inferences to what was done, and therefore to what one is responsible for.
BIBLIOGRAPHY


