FREEDOM AND RESPONSIBILITY

by

BRUCE BAUGH

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ARTS
in THE FACULTY OF GRADUATE STUDIES
in the Department of

Philosophy

We accept this thesis as conforming
to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA
August 1978

© Bruce Baugh, 1978
In presenting this thesis in partial fulfillment 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 his representative. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Bruce Baugh
ABSTRACT

This thesis shows the ways in which the concepts of freedom and responsibility are related, and how indeed they illuminate each other. In Part One, it is shown that both are based on a concept of action, and it is thus with an analysis of action that a theory of freedom and responsibility must begin. Actions are first differentiated from events, so that conditions which must obtain from an event to be an action are specified. The concept of responsibility may then be used to illuminate action by showing how excuses indicate ways in which actions can fail. From this analysis, an analysis of action in the full sense emerges, namely, that an action in the full or unqualified sense is that to which no excuses are applicable. Action in the full sense is thus linked to responsibility in the full sense. The analysis of action shows that the breakdown of an action is the loss of control over its effects, and action in the full sense thus obtains where no breakdown occurs. Conscious control over an action is the control of an action's effects, which is the realization of intentions, and the control of intentions, which is what may be analyzed as rationality. Conscious control over an action, or agency, constitutes freedom on the plane of individual action. Thus, from the concept of responsibility emerges a concept of action and of agency which indicates what freedom is. Yet, it is the actual structure of action upon which the action of responsibility rests.

The theory of freedom is defined in Part Two against the incompatibilist position that if determinism is true, neither
freedom nor responsibility exist. It is shown that causal determinism does not rule out actions' being free in the sense required for an individual to be responsible for them as a theory of action shows that it is not an action's being caused but the nature of its causes which makes it free or unfree. If the action is caused so that it is in the conscious control of the agent, it is free. The rest of Part Two examines moral practices such as praise and blame in light of the limits determinism places on them. It is necessary to show what rational or justifiable grounds there could be for practices such as praising and blaming in any theory of freedom and responsibility.

Part Three shows that agency, or control over an action, is extendable over the values upon which actions are based. Control over values is achieved by the individual consciously choosing values in awareness of being responsible for those choices and values. What the Existentialists call "Authenticity" is thus a fuller degree of freedom and of agency. This analysis of authenticity does not focus on how authenticity is a response to a value question posed by nihilism, but on how authenticity is an extension of our regular concepts of freedom and responsibility. It is shown that authenticity, when it is accompanied by full agency (as that notion is developed in Part One) is freedom and responsibility in the highest degree.
<table>
<thead>
<tr>
<th>FREEDOM AND RESPONSIBILITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
</tr>
<tr>
<td>Part One: Action and Responsibility</td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
</tr>
<tr>
<td>Causality and action</td>
</tr>
<tr>
<td>Intention and effect</td>
</tr>
<tr>
<td>The structure of action</td>
</tr>
<tr>
<td>Rationality</td>
</tr>
<tr>
<td>Involuntary actions</td>
</tr>
<tr>
<td>Conclusion</td>
</tr>
<tr>
<td>Part Two: Determinism and Freedom</td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
</tr>
<tr>
<td>Defining the problem</td>
</tr>
<tr>
<td>Capacities and freedom</td>
</tr>
<tr>
<td>Incompatibilist objections: Campbell and Hospers</td>
</tr>
<tr>
<td>Praise and blame and responsibility</td>
</tr>
<tr>
<td>Character and moral worth</td>
</tr>
<tr>
<td>Praise and blame as attitudes and as actions</td>
</tr>
<tr>
<td>Praise and blame and agency</td>
</tr>
<tr>
<td>Reactive attitudes</td>
</tr>
<tr>
<td>Part Three: Authenticity and Freedom</td>
</tr>
<tr>
<td><strong>Introduction</strong></td>
</tr>
<tr>
<td>Degrees of freedom</td>
</tr>
<tr>
<td>Value and choice</td>
</tr>
<tr>
<td>Sartrean freedom</td>
</tr>
<tr>
<td>Authenticity as freedom</td>
</tr>
<tr>
<td>Conclusion</td>
</tr>
<tr>
<td>Conclusion</td>
</tr>
<tr>
<td>Notes</td>
</tr>
<tr>
<td>Bibliography</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENT

I would like to acknowledge the help and influence of Prof. S.C. Coval, who introduced me to the approach taken in this thesis, particularly in Part One. The method used in Part One is borrowed freely from Professor Coval's unpublished paper, "The Concept of Action."

I would also like to thank my advisor, Earl Winkler, for his guidance, support and helpful comments and criticism.

My thanks also to Paisley Woodward for encouragement and fruitful discussion.

B.B.
INTRODUCTION

Freedom and responsibility are concepts that are interconnected in such a way as to illuminate each other. Both are based on the concept of action. Any theory of freedom and responsibility must then begin with an analysis of action. In the first place, it is necessary to differentiate actions from other sorts of events, and so it is necessary to show the conditions which must obtain for an event to be an action. Once these have been established, the concept of responsibility may be used to reveal the structure of action by pointing out how excuses indicate ways in which actions fail. From this analysis a description of action in the full sense emerges, namely, that action in the full or unqualified sense is action to which no excuses are applicable. Action in the full sense is thus linked to responsibility in the full sense. The notion of action is linked to that of agency, and from the analysis of action a theory of agency emerges which holds that full agency is the possession of conscious control over an action. The analysis of action shows that the breakdown of an action is the loss of control over its effects, and action in the full sense, and hence full agency, thus obtains where no breakdown occurs. Conscious control over an action is the control of an action's effects, which is the realization of intentions, and the control of intentions, which is what may be analyzed as rationality. Conscious control over an action, or agency,
is what constitutes freedom on the plane of individual action. Thus, from the concept of responsibility emerges a concept of action and of agency which indicates what freedom is. Yet, it is the actual structure of action upon which the notion of responsibility and its application in practice rests. The notions of freedom and responsibility are thus interdependent and have as their foundation a theory of action, which is itself illuminated by the concept of responsibility.

The theory of freedom developed on the basis of a theory of agency can be defended against the objections of determinists or incompatibilists, who hold that neither freedom nor responsibility can exist if determinism is true. It can be shown that, whether or not determinism is in fact true, the causal determination of actions does not rule out those actions being free in the sense required for an individual to be responsible for them. On the basis of the theory of action, one can show that it is not an action's being caused, but the nature of its causes which makes it free or unfree. If the action is caused in such a way as to be within the conscious control of the agent, then it is free. Determinism does place limits on the concept of freedom, however. Specifically, it rules out a notion of freedom which is the basis of some of our moral practices, notably those of praising and blaming. It is necessary to outline what rational or justifiable grounds there could be for practices such as praising and blaming if determinism is true in any theory of freedom and responsibility.
Agency, or control over an action, is extendable over the values upon which actions are based. Control over values is achieved by the individual consciously choosing values and being consciously aware of being responsible for those choices and those values. What the Existentialists call "authenticity" is thus a fuller degree of freedom and of agency. As such, it entails the highest degree of responsibility for actions which an individual is capable of assuming. In the context of a theory of freedom and responsibility, an analysis of authenticity focusses not on how authenticity is a response to a value question posed by nihilism, but on how authenticity is an extension of our regular concepts of freedom and responsibility, and shows that authenticity, when it is accompanied by full agency, is freedom and responsibility in the highest degree.

The notions of freedom and responsibility developed in this analysis are applicable to primarily individual actions, and so it is with a theory of action that we begin and with a theory of the individual's own responsibility for actions that we close.
PART ONE

ACTION AND RESPONSIBILITY

1. Introduction

The concept of responsibility usually discussed in the determinist/libertarian debate is one which concerns action. In order to get clear on what it is to be responsible for an action, one must first understand what it is for someone to act. Once a basic notion of action is arrived at, one can then develop a concept of agency, which would refer to the powers and abilities of individuals to act in various circumstances, and where those powers may be diminished. The analysis of the concept of agency will refer to the various ways in which actions can fail, or break down, and thus shows how agency can be diminished. This analysis illuminates the central core of the concept of action, indicating the minimum conditions necessary for agency. By seeing how actions fail and how agency may be diminished, a notion of what it is to have full agency will emerge. Part of the notion of full agency is the concept of rationality, and it is this concept that lies at the basis of social rules and practices, including the institution of responsibility. From the concept of agency, one will then be able to develop the concepts of freedom and responsibility in action.
2. Causality and action

The first point to consider is how an action differs from other natural events. If determinism is true, then actions, like events, have causes. The difference between actions and other natural events is that the system of causes of an action must be in some way internal to the agent in a way that a system of causes of other events is not internal to them. It will be shown that not all bodily movements are actions, for not all bodily movements have the right sort of causes to be called actions.

It seems that an action is an event caused by the agent. To make it clear in what way it is caused by the agent, it will help to consider the view that there are some actions ("basic actions") which the agent does not cause. Danto argues¹ that although there is a sense of "cause" such that it may be said that an agent "causes" an action, this is a different sense of "cause" than the one used to describe an event being "caused" by the agent. The sense of "cause" in which the agent causes an event (which is not itself an action of that agent) entails that the event was caused by an action of the agent. For example, to say an agent caused the rolling of the rock implies that the agent performed an action (eg. pushing) which caused the rock to roll. (This implication seems to hold, however, only where an agent may be held responsible for an event. Helen of Troy may have been the cause of the Trojan War, but we would not hold her responsible for that event in that it was her beauty, and not her actions, which
caused it, and so she did nothing to cause it.)

There are at least some actions which are not caused by the agent in the way an event may be caused by an agent (which is through some action of that agent). That is, there are some actions for which the agent performs no additional or separate action in order to bring that action about. For example, in order to move my arm, I need not perform any previous action. I did not cause my arm to move by first trying to move it, or by contracting muscles or firing neurons, for the firing of neurons and the contracting of muscles is the same event as the arm's being moved by the agent (and so is the same action), and "trying to move an arm" implies only that the action may not have been successful, but if it were successful the trying and the actual movement would be the same action under different descriptions, and so "trying to move an arm" implies no action antecedent to the actual movement of the arm. The movement of an arm would differ from the agent's trying to move it, or the agent firing certain neurons and contracting certain muscles, only if the source of movement is external to the agent (it could be a pulley, or someone else moving the arm), in which case the movement is not an action of the person whose arm moves. So there are some actions, according to this argument, which are not caused by the agent in that there are some actions for which the agent does not do anything in order to bring it about, and to say an agent causes something is to say the agent has done something to cause it.
The point at which there is no previous action that is necessary to produce the action is the point where what takes place is not a deliberate exercise of power by the agent, but some process internal to the agent over which the agent has no control. The beating of a heart is not an action, but theoretically it could be, if one learned how to control it. Similarly, some brain processes are usually simply internal events, not actions, but if one could control them, then one would be acting when one produced such events. Through biofeedback techniques, some people can learn to produce certain brain events, in which case those events become actions, as they become subject to the conscious control of the agent. (Returning to the Helen of Troy example, it is clear that Helen's beauty was not an action in that she had no conscious control over it. Thus though she caused the Trojan War through her beauty, she performed no action, and so could not be held responsible for the war.) Events not subject to the conscious control of the agent are not actions, but mere events (unless they are subject to the conscious control of another agent, in which case they are actions of that agent). It seems that there must be some point where the cause of an action is not subject to one's deliberate control, so there is some point at which an action is caused by an event which is not an action. If the point at which an event becomes subject to the conscious control is the point at which it becomes an action, then an action is an event within the conscious control of the agent. Actions not caused by any
previous action are what Danto calls "basic actions," and basic actions, since they are caused by events over which the agent has no control, are not caused by the agent.

Yet there is another sense of "cause" such that one could say that the agent "caused" a basic action. Basic actions are caused just as other physical events are caused, but possess a unique and differentiating characteristic. For example, the cause of a basic action may be the firing of certain neurons. The firing of those neurons does not seem to involve any intentionality or deliberate agency, and is thus like other natural events which are not actions. In that the agent has no power over the firing of those neurons, and thus performs no action which produces such firing or the basic action which results from it, the agent cannot "cause" basic actions. But in so far as the cause of the action is internal to the agent, then the agent is the cause of the action in the sense that those causes of the action internal to the agent are necessary conditions of the action's being performed. Even though the agent has not done anything to produce the action, the system of causes within the agent which produces the action is (in a sense) part of the agent, and therefore one may state that if an event is an action, it must be caused by the agent (in the second sense of cause just indicated). If it were not, then there would be no difference between actions, basic or otherwise, and other events. It is the origin of the action in a system of causes internal to the agent (and thus, the origin of the action in the agent) which makes the action an
action of that agent. This is only to say that for an event to be an action, the system of causes which produced it must be at least partially internal to the agent. This is a necessary, and by no means sufficient, condition of an event's being an action, as there are some events produced by causes internal to the agent which are not actions (as will be shown later).

Can machines act? Apparently their movements and operations are produced by causes which are to a large degree internal to themselves. It seems reasonable to assume that to the degree to which a machine performs operations autonomously (that is, without being directly controlled by someone or something else), it may be said to be acting. An internal combustion engine does not act, but a robot could. If a machine operated autonomously, it would have to be self-motivated, and therefore its operations would be intentional (a deliberate exercise of power for its own reasons or purposes), and therefore actions, for which it could be held responsible. Thus, a machine could not act unless it had some intentions, and it could not have intentions without some sort of concept of self (though the concept need not be the self-conscious awareness of self that humans possess). Machines could act, therefore, and so this possibility is no objection to the point that actions are produced by a system of causes internal to the agent, whether the agent is human or mechanical.

Of course, there are some actions which are said to be caused by other actions. (We are not speaking here of cases
where the action of one agent is said to be the cause of another action of the same agent.) For example, in order to move a rock one must first push it. There seem to be two actions here, the pushing and the moving. But the pushing and the moving of a rock are one movement, one effort and one action. The same physical event (neurons firing, muscles contracting, force being exerted) takes place. To try to move the rock is to push it, and if the attempt is successful, no action in addition to the pushing is needed to accomplish it. The apparent difference between the two actions (the moving and the pushing) is merely that one is described with reference to the effect of the action (the movement of the rock) and the other is not.

There is a confusion in saying that an action can cause another action which may be brought out if one notes cases where an agent performs an action that produces another bodily movement that is not an action. If an external force (such as a sling operated by a mechanical pulley) causes someone's arm to move, that movement is not an action, as the system of causes of the event is external to the agent. But if someone uses one arm to move the other (as when, for example, the other arm is paralyzed), is the movement of the arm being moved an action? In one sense, the system of causes of the movement is internal to the agent, and so it would seem that the movement of the arm caused by the movement of the other arm is an action caused by another, previous action. However, closer analysis of this example shows this not to be the case. The arm being moved by the other arm is being moved not simply
by a system of causes internal to the agent, but by the external force generated by the system of causes internal to the agent which is the movement of the other arm. In this respect, the movement of the paralyzed or inactive arm is like the movement of a rock that is pushed; the agent moved the inactive arm in the same way that the agent would move a rock, and not in the way that the agent would move a normally functioning arm. The movement of the arm which causes the other (inactive) arm to move is caused by nothing except factors internal to the agent. That is, if one traces the causal chain of the movement back to the moment before the movement took place, the causes would all be internal to the agent (though one could, of course, trace the causal chain back further to a point at which at least some of the causes, as, for example, an object of perception, are external). It is not caused by some previous movement or action. The arm being moved by that functioning arm is an action of the agent's only in the way that a stone being moved is an action; it is an action in that it is a movement caused by an action (and so is that action described with reference to its effects). But the movement of the inactive arm in and of itself is not an action, in the same way that a stone rolling is not in and of itself an action. Rather, it is the effect of an action, for the arm being moved is merely the object and not the instrument of the action in such a case (since one could not intentionally move it without performing some separate and previous action). The arm which is inactive or
paralyzed does not participate in the action any more than
does a stone that is rolled. Because of this passivity in
the arm, it makes no difference whether the force that moved
it was produced by the agent or not, so long as it is external
to it in the way that has been described, in which case it is
not an action in and of itself but the effect of an action.
(The movement of an artificial limb is not in and of itself
an action; it is only an action when it is an effect of some
previous or separate action, such as the movement of a partial
limb or of certain muscles. The movement of the partial
limb or of the muscles and of the artificial limb are one
action and one movement, as the movement of the artificial
limb is an effect of an action, not a second action produced
by the movement of the partial real limb or of certain muscles.
The movement of a natural and functioning limb, by contrast,
requires no previous action, and can only be an action of
the person of whom it is a part if no previous action, such
as the action of someone else, produced it, and thus must
result from the system of causes internal to the agent if it
is to be an action of that agent.) Just as we can say "the
agent rolled the stone," we can say "the agent moved his or
her paralyzed arm (by lifting it with the other arm)," which
is simply a way of describing an action of the functioning arm
doing the lifting in terms of its effects. The paralyzed
arm does not act, and its movement is not an action. (It is
impossible for an agent to simply move a paralyzed arm, for
a paralyzed arm is one that cannot be moved, unless it is
moved by another arm, a machine, or someone else.) There are not two actions (the movement of the functioning arm and that of the inactive arm) such that the one is the cause of the other. It is clear from this analysis that not only must an event be produced by a system of causes internal to the agent in order to be an action, it is necessary to distinguish between actions and their effects.

3. Intention and effect

The first reason for making the distinction between actions and their effects is that it follows from the description of an action being the result of a system of causes internal to the agent. We have just seen that this description shows that no action is the cause of another action. This applies to the actions of different persons as well as to actions performed by the same agent. If the action of an agent is genuine, then what the agent does may be caused by someone else's action only in an indirect way. For the action of another can be the "cause" of an action only in the sense that the action of the agent is a response to the action of the other (just as an action can be a response to some other factor beyond the agent). This does not mean the other person caused the action (or that other elements of the environment caused the action), for that would mean that the action was merely the effect of the action of another, and not the action of the agent at all. (If the action was caused by other factors in the environment than the action of another it still would not be an action, for
It would be a mere event as it would not be produced by a system of causes internal to the agent.) Hence actions cannot have as their complete cause another action (or event); they must result from the system of causes internal to the agent. This system of causes may itself be affected by factors both within and without the agent (such as the actions of others), but if it is not operative, the agent does not perform an action.

There is an additional, pragmatic reason for distinguishing actions from their effects, which is that to say someone has acted is to ascribe responsibility for the effects of that action to that person. There may be cases where such a description is inappropriate unless qualified by some adverb of excuse, as there are cases where a person has acted but where we would not want to hold the person responsible for the effects of the action. In such cases, something has gone wrong with the action, so that the intended effect is not achieved. So while the agent has acted, it is wrong to describe the action in those cases in terms of the effect of the action, for to do so is to ascribe (by implication) responsibility to the agent for the effect. In such cases, where the effect achieved was not that which was intended, any description of the action in terms of the effect must be qualified somehow to indicate that the effect was unintentional, and therefore that the agent may have diminished or have no responsibility for it. In law, one means of defense is to admit performing an action, but to claim that the description
of the action must be qualified (eg. with reference to the circumstances, the agent's state of mind, etc.), thus diminishing or eliminating the agent's responsibility for the action. Thus, the system of causes within the agent may produce an action which goes awry for one reason or another, and so while the agent has acted, the only unqualified description of the action that will be correct will be that which refers to the action itself (or what Danto calls the basic action), and not to the effects of the action.

The objection may be raised here that responsibility is diminished or eliminated in cases other than those where the effect of the action was unforeseen or unintentional, as for example with involuntary or coerced behaviour. However, it will be shown that involuntary or coerced behaviour is usually behaviour for which the agent is held responsible, but of which we would not, considering the circumstances, disapprove. Thus, it is not excused, but is in a sense justified or condoned. From this it will be argued that where the agent is rational and there is no breakdown between the intention and the effect, then there is full responsibility. For simplicity, mistake of fact and irrational belief will be treated as a breakdown of agency, as these are cases where the intention itself breaks down as it is not compatible with the other intentions of the agent, but this incompatibility is hidden from the agent due to a mistaken or irrational belief.

("Intention" is here used to mean the directing of oneself toward an end, and so is coincidental with the postulation and
pursuit of an end, as distinct from a desire, which is merely a tendency to postulate and pursue an end. To intend something is then not only to desire it, but to direct oneself toward it as an end, or to pursue it, and so it is not a psychological state or attitude which precedes action but is a component of ends or of goals.)

The link between the ascription of responsibility and that of agency is so strong that it led H.L.A. Hart to claim that the concept of action cannot be explained by descriptive statements. Rather, he says, the concept must refer to ascriptive statements about responsibility. "Our concept of action ... is a social concept and logically dependent on certain rules of conduct." This is to put the cart before the horse. The determination of actions, or of how an action should be properly described, is not independent of normative considerations about under which conditions and in what way one should hold people responsible. So Hart is right in that respect, but is also misleading. Because of the conceptual connection between the concept of action and that of responsibility, we need a concept of responsibility to illuminate why certain actions are excused. What is finally explained, however, is the concept of action itself (including how actions break down), which is the basis of the concept of responsibility. The reason why some actions are excused must refer to the concept of action. The evaluative or normative aspect of the concept of action relies not so much on rules of conduct but on some concept of
agency, which focuses on how agency may be diminished in certain circumstances. The evaluative component of the concept of agency is some concept of intentionality, which indicates what criteria there are for ascribing agency and hence responsibility. It is thus a normative concept of intentionality, not rules of conduct, which indicates how and when standard deviations from standard action and agency are excusable. The normative core of the concept of intentionality is a concept of rationality which provides a rationale or basis for rules of conduct and for the concept of responsibility, as will be shown later. The phenomenal basis for the concepts of action and responsibility is thus the actual structure of action, and the normative basis for the concept of responsibility is the concept of rationality as it applies to social practices. To justify excusing actions on the basis of the concept of responsibility would be to engage in a *petitio principi*. Rules of conduct must be logically dependent on what we know about how people act in order for the rules to be efficacious. Unless someone knows enough about action to be able to demonstrate why persons are held responsible in some situations and not in others, or why some people are never held responsible, it will be impossible to justify the social practice of holding persons responsible. If the practice is to be justified, it must make reference to the facts of the world which make the practice efficacious, and since the practice governs actions and refers to them, it is natural to suppose that the facts in question concern the concept of action. To
say that someone performed an action without qualifying that statement implies that the person was responsible for the effects of the action, but it is not the rules of conduct alone which provide the justification for saying whether or not the person is responsible for the action, for what is at issue is whether or not the person acted in conformity to those rules of conduct, and the answer to that question will depend upon an examination of the person's actions, and not just the rules of conduct. In order to ascribe responsibility, one must first be able to decide whether and in what manner the person acted. If whether the person is responsible depends on the extent to which the person acted in the standard (i.e. unqualified) sense of action, then it is the concept of action which determines when a person has acted in the unqualified sense of action, and which therefore determines what it is to be an agent and to be responsible.

4. The structure of action

Since Aristotle, much has been made of the fact that the full concept of agency is determined negatively. That is, the standard concept of action is that action where all excusing conditions are absent. But grounds for exception from responsibility (conditions of excuse) only diminish agency, they do not remove it, and so a central concept of action (the "core concept") remains after all the aspects of agency cancellable by conditions of excuse have been so cancelled. Adverbs of excuse do not deny agency, but they do deny agency in the standard sense. Standard agency is then that core
of action which remains after all aspects cancellable by adverbs of excuse have been cancelled, plus all those cancellable aspects. (Add to what remains of the whole what was taken away from the whole and the result is the complete whole, but understood now in terms of its parts and its principle of unification.) The standard case of action is thus one where no excusing conditions are applicable. In that standard agency is the absence of excusing conditions, it is defined negatively, but in that it makes reference to the core of action which is not cancellable by excusing conditions, it is not. The core of action is not cancellable in that its absence would entail that the object of investigation is not an action, and it constitutes the essence of action in that a description of it is a specification of the conditions necessary for something to be an action. The way in which the parts of standard action which are cancellable are related to standard action indicates the structure of action. The apparent heterogeneity of excuses and of the conditions to which they refer obscures the fact that they have in common the quality of pointing out how an action can go wrong. That is, they point out standard deviations from the standard or paradigm case of agency. 7 It can be shown that an action goes wrong when the system of causes within the agent that produced it is defective or is made fully or partially inoperative by some external condition. When an action produces an undesired or unintentional result, something of that sort can be seen to occur, if one includes such things as reasons and intentions
within the set of causes internal to the agent which produces actions. (Though an intention is a component of action, it may be said also to be a cause of action in that it is a necessary condition of an action's being performed, since actions must have some goal or object, but it does not follow from this that an intention is a psychological state which precedes action. Intentions precede the completion of an action, but not the commencement of an action's execution, as to embark on an action by directing oneself toward a goal or an object and to have an intention is the same thing.)

There is little disagreement as to what kinds of conditions are taken to be conditions of excuse. There are words in our language (adverbs of excuse) which refer to standard ways in which the system of causes of an action internal to the agent may either break down or be made inoperative. By seeing how an action may go wrong, we get a sense of what a standard action is, as well as the minimal concept of action that remains in the presence of excusing conditions. This is because the cancellation of certain features of action by excuses indicate the way in which the deviant case (the excused action) is related to the standard one.⁸

A person is not held fully responsible for the effects of an action unless those effects were intended. Even where the agent was negligent, the agent is responsible for the negligence and the effect of it, but the agent is not held responsible to the same extent that a person intending that effect would be held responsible.
There are various ways in which the effectiveness of the system of causes internal to the agent may be overridden by the presence of some factor that produces an effect other than the one the agent intended. It is not that the system is completely overridden, for it it were the action would not have been produced by the system of causes internal to the agent, in which case it would not be an action of that agent (although it could be the effect of the action of another agent) and the agent would not be responsible for it at all.

What is overridden here is the standard system of causes within the agent, and so the reasons for ascribing standard agency and full responsibility are overridden. If in the course of an action an unforeseen event occurs which alters the effect of the action so that it is no longer the effect intended (and it may happen that the effect produced is contrary to that intended), this is an instance of the system of causes of action within the agent being partially overridden by some other causal factor, in that part of the system of causes (or, one may say, the whole of the standard system of causes) is made inoperative. The intentions (which are part of the system of causes) are not realized, because they are overridden by some other, more powerful factor. This is not to say that intentions did not operate as a causal factor (for if this were so, there would be no deliberate exercise of power and hence no action), but only that some other causal factor caused those intentions to remain unrealized, or defeated. If that causal factor is an unforeseen event, we are liable to say that the
consequences of the action were accidental, and we would diminish or eliminate responsibility for that action to the extent that the factor affecting the outcome of the action (in this case, the factor is an event) could not have been reasonably foreseen. Standard agency, and full responsibility, occurs when a person is able to foresee all relevant circumstances or causes of the actual effect of the action, including events that may occur between the initial operation (e.g., a bodily movement) and the effect. In such cases, the complete causal mechanism internal to the agent (including intentions, which are preserved due to the presence of foresight) is the chief causal factor producing the effect (in that the agent is able to determine that effect through action) and is not overridden by other causal factors. Adverbs such as "accidentally" cancel this aspect of standard agency, and any description of an action where that aspect (i.e., of foresight) is missing that is not modified by a qualifying adverb such as "accidentally" would be inaccurate, as the unmodified description implies standard agency and full responsibility.

One may note that the fact that "accidentally" may refer to other instances of an action producing an unintended effect than the sort just outlined does not concern us here. What it is important to notice is the different stages of an action where that action can go wrong. There are many ways for intentions to be overridden and made inoperative. One way is for an unforeseen event to alter the outcome of the action, as in Austin's example of the neighbor's donkey moving in front
of a target and being accidentally shot. The movement of
the donkey was a causal factor more powerful than the agent's
intentions and efforts to realize those intentions, and so
overrode the system of causes of the action internal to the
agent by causing those intentions to go unrealized. The
intentions were made inoperative by the movement of the
donkey in that they no longer determined the effect or con­
sequence of the action, whereas the movement of the donkey
did. The presence of foresight (in the system of causes internal
to the agent) would have allowed the intentions to retain their
causal efficacy by eliminating the donkey's movement as a
causal factor.

An action can go wrong when in the normal performance of
the act an additional effect consequent to the effect intended
or resulting from the action as an unintentional side-effect
is produced. This is an instance of the system of causes of
an action internal to the agent being overridden by factors
present in the situation at the time of the performance of
the act. Although the system of causes is effective insofar
as it produces the action intended, it is overridden in that
it also produces undesired or unintentional results (which
mean that the intentions which are part of the system of
causes are overridden in that there are some effects of the
action which they do not determine). This differs from the
way an action can go wrong already described under the rubric
of "accidents" in that no unforseen event causes the uninten­
tional effect of the action. For this reason, this sort of
unintentional effect (due to inadvertance) is less excusable, for it results from the agent not paying attention to all the factors of the situation in which the act was committed. The agent was not responsible to the extent that the effect was unintentional, but is not excusable to the extent that had proper attention been paid to all the circumstances of the act, the unintentional (and perhaps undesired) effect could have been avoided. What will count as "proper attention" will depend on the ability of the agent to have noticed details of the situation given the agent's epistemic capacities and the nature of the circumstances. The knowledge the agent has of that type of circumstance would also be relevant, for if one could not reasonably have expected the agent to know about the factor that produced the inadvertant consequence of the action, then one could not reasonably hold the agent responsible for that consequence. For example, a child in a new situation about which he or she has no knowledge (and could not reasonably be expected to have knowledge due to the lack of opportunity to have obtained it) would not be held responsible for the inadvertant consequences of its actions in that situation. Situations where the agent's epistemic capacities are diminished or made temporarily inoperative due to circumstances, or where the agent is lacking in such capacities, will be ones where the agent will not be held responsible for any unintentional side-effects or additional effects resulting from inadvertance, insofar as the inadvertant action was due to the absence of dimunition of the agent's capacities. (A
further discussion of the capacities of the agent in relation to responsibility comes later in the paper.) A fully responsible agent would be one both able to foresee all the consequences of the action and who had in fact foreseen all the consequences. (An agent able to foresee the consequences but who did not would either be excusable or negligent, depending on the circumstances and the kind of inadvertence, but in neither of these cases would the agent be fully responsible for the unintentional and inadvertant effects of the action.) The standard case of action would then be one where all the consequences of the action were foreseen, and it is this feature of standard action which is cancellable by adverbs of excuse such as "inadvertantly." From what has been said already, it is clear that a case of an action where some effects were not foreseen but resulted anyway as a matter of inadvertance in the normal performance of an act that is described without using qualifying adverbs of excuse is described incorrectly, for such a description ascribes more responsibility to the agent than we would want to in this case.

There are cases where the system of causes internal to the agent may produce an action which is itself not intended. This kind of case differs from the others in that it is described with reference to something going wrong in the initial or basic action, rather than to the action described in terms of its consequences. For example, the adverb "carelessly," refers to inattention on the part of the agent
to the manner in which the action is performed, and not to the unforseen consequence of the action (as in "she moved her arm carelessly"). As with other examples mentioned so far, it is entirely irrelevant that the adverb of excuse, in this case, "carelessly," may have other meanings in other contexts. One can perform a "careless" act deliberately and intentionally, in which case the adverb refers to the fact that the person did not care about certain values or goals we generally feel that people should care about. For example, one can toss garbage down a ravine "carelessly," and what is meant here is that one does not care about the harmful consequences of one's action. But used in that sense, the adverb is no longer one of excuse. One can also use "carelessly" to refer to an action in terms of its effects, as in "she carelessly knocked over the pitcher," but what we are concerned with here is in showing the different ways in which actions can fail, and not in providing definitions for words in the English language. Thus, it is nothing to the point that "carelessly" may be used in other ways that it is used here, for the use of it here is merely as a label for a certain way in which actions can fail and agency can be diminished.

So to return to our example of acting carelessly, since the basic or initial movement is forseen and intentional (when described without reference to its effects), the carelessness lies in not paying attention to the performance of the action, and not in inattention to the present situation (as in inadvertance) or in the inability to forsee future events which affect the outcome of the action (as in accidents). An agent,
through carelessness, may then perform an act $x_1$ while performing act $x$ by not paying attention to the manner of performance of act $x$. Thus, an intentional and forseen initial action may go awry and become an unintentional action. Adverbs such as "carelessly" cancel the feature of standard action which is the agent's exercising a certain amount of care and attention in the performance of the act.

Carelessness is not a very powerful excuse, as it is a form of negligence, rather than simply ignorance of the facts, and could have been prevented by the agent. It is excusing only insofar as it indicates that a result of the action other than that intended may have been produced due to inattention to the manner of performance of the action. The agent is not responsible for the consequence or effect that resulted from negligence insofar as it was not intended, but is responsible for it insofar as it resulted from negligence which could have been prevented. Carelessness is not a powerful excuse because the care one takes in the performance of an action is something over which one usually has complete control, in contrast to aspects of the environment or future events (which affect the outcome of an action). There may be some cases where carelessness is excusable because the agent did not have full power over the manner in which the action was performed, as, for example, when an agent is distracted and so prevented from paying sufficient attention to the manner of performance of the action. Some distractions would be more legitimate excuses than others, as being distracted may result simply
from a lack of effort to pay attention to what one is doing (in which case the excuse is not very powerful), but in other cases, there are some events (such as a person crying for help) which we could not reasonably expect someone not be distracted by (in which case the excuse is quite powerful). Full agency thus involves having the opportunity to pay attention to the performance of an action and actualizing that opportunity. The carelessness which results from the lack of opportunity to take sufficient care (due, for example, to the presence of a compelling distraction) would be more excusable than that which results from the failure to actualize that opportunity, as the former indicates a breakdown in the system of causes that produces an action over which the agent has less, if any, control than over the latter, which excuses only because it indicates that inattention to the manner of performance of the act occurred which may have led to effects which were not intended (i.e. the intentions were not realized). The breakdown of action in the latter case is due to an event which the agent has the power to control and which therefore could have been prevented by the agent, and so is not grounds for a very powerful excuse.

A breakdown in action can occur when the agent's beliefs about the facts of the situation in which the action takes place are mistaken. The effect of the action is not that intended because of the epistemic state of the agent. As the excuse "mistakenly" refers to the agent's epistemic state, it may be assumed that in the standard case of action the agent's
epistemic state is such that the agent holds no false beliefs about the facts of the situation which affect the outcome of the action, and so holds no false beliefs about the likely affects of the action. Standard action and full agency may obtain when one holds false beliefs which could affect the outcome of the action so long as those beliefs do not in fact affect the outcome, for then, despite the presence of false beliefs (even if the beliefs are about or relevant to the situation in which the action takes place), the system of causes of the action internal to the agent would have produced the action without having been overridden at any point, for the agent's intentions would be realized, and so the agent would be in possession of full agency, and consequently would be fully responsible for the action, all other things being equal.

5. **Rationality**

So far, we have assumed that the agent acts with a view to achieving some end which the agent wants, and that an action may go wrong when what is intended is not achieved. However, there may be cases where an agent may be mistaken about what he or she wants. This mistake could result from an inability to see the relation between one goal and another (so that a secondary goal is acted upon at the expense of the primary goal, unbeknownst to the agent), or from a mistake as to what means will best secure the desired end (which results in that end not being achieved). These mistakes about what the agent really wants may result in the agent never attaining those
goals which are most important to that agent. Moreover, the mistake may be so deeply rooted that the agent is unable to apply new and relevant information to the problem, and so repeatedly makes the same kind of mistakes. In such cases, the agent acts not simply mistakenly, but irrationally.

The notion of rational action put forward here is a critical one, for it entails that only those actions motivated by good reasons (i.e., reasons which meet certain evaluative criteria) are rational. The concept of rational action has also been analyzed as an empirical concept, notably by Hempel. The empirical hypothesis is that those actions which can be explained with reference to the reasons which produce them are rational. But as Hempel fears might be the case, such a concept of rationality is otiose. What Hempel's investigation of the empirical hypothesis of rational action reveals is that all intentional actions are motivated. To put it another way, there are always reasons for actions which we are conscious of performing and which have some purpose. As it is pretty clearly not the case that all intentional or purposive actions are rational, something must be wrong with the account of rational action proposed by the empirical hypothesis. For this reason, only a critical concept of rationality may prove useful. The critical concept judges an action as rational by evaluating the agent's reasons for the action and their relation to the action produced. This evaluation attempts to determine whether or not the reasons for the action were good, and the causal efficacy of those reasons in producing the desired outcome.
The reason that Hempel is inclined to consider an empirical account of rational action is that his account of the critical concept is mistaken. According to Hempel, the critical analysis of action judges as rational actions which, in the light of the agent's objectives and beliefs, are appropriate means to attaining the end sought. The rationality of an action thus depends on (1) the objectives the action is meant to achieve and (2) the relevant empirical information available to the agent at the time of the decision. Hempel contends that an action is rational if it offers the optimal prospects of achieving its objectives.

As Hempel points out, in assessing the rationality of an action one must consider only the facts (concerning which means may best achieve the agent's objective) which are available to the agent, and not all the facts of the situation which would be relevant to the agent's decision to act in one way or another. The relevant facts are those which pertain to the circumstances in which the action is to be performed, the different means which in those circumstances are likely to achieve that end, the side-effects (which are those effects of the action which are additional to the effect intended but are secondary to some primary goal or intention) of the action (considered as a means) and (a factor which Hempel omits) the state of the agent. It would be incorrect to judge an action as irrational if it failed to achieve optimal results due to the unavailability to the agent of some relevant facts. In this we can agree with Hempel.
Hempel goes on to say that one must assess the rationality of an action in the light of the agent's beliefs, but the basis of those beliefs he claims is not important for the analysis. Many apparently irrational actions, the argument runs, are rational given the agent's beliefs, as the actions proceed from those beliefs. But if one is to declare with Hempel that an action may be rational given the beliefs of the agent, then any intentional action will be counted as rational. An intentional action is purposive: in other words, the agent has reasons for acting (some of which would be beliefs). Hempel is right in saying that to show that an agent acted irrationally one would not have to show precisely what the grounds that produced that belief are, but he fails to see that this is because one would have to show merely that the grounds of the beliefs are inadequate. It is not necessary to know why or on what grounds Jill believes she will experience bad luck if she walks under a ladder in order to show that such a belief is irrational; it is sufficient to show that there is no good evidence which could support such a belief.

We can see now why Hempel's account of rational action is neither full nor accurate: he neglects to point out that for an action to be rational, the beliefs on which it is based must also be rational. In other words, a full critical account of rational action must take into account not only the beliefs of the agent, but the basis of those beliefs. A rational action would then be one produced by good reasons, where good reasons are beliefs based on evidence available to the agent (about both means and ends). There are cases of irrational belief,
that is, cases where the agent holds a belief not supportable
(or even made probable) by the available evidence; and indeed,
in some cases the beliefs run counter to the available evidence.
It seems plausible to say that any action based on an irrational
belief is itself irrational.

Reasons similar to those outlined above indicate that the
objectives of an action (including side-effects, when these are
taken into account) and the norms constraining the choice of
means by which the objectives may be achieved may also be
classified as rational or irrational. Any objective is a
product not of the agent's needs, but of what the agent believes
those needs to be. There are cases where the agent may act
contrary to his or her own interests. This in itself may not
be irrational; it would be foolish to judge all mistakes or
miscalculations about what is in one's interests as indications
of irrationality. It is an indication of irrationality only
where the beliefs upon which the objectives and norms of action
of the agent are based conflict with, or are not supported by,
the evidence available to the agent (eg., concerning the
agent's own interests). In such cases, the objectives or
norms of action are irrational. (Of course, if the discrepancy
is slight the false belief may be more on the order of a mistake
than an irrationality, but the greater the severity of the
discrepancy, the greater grounds there are for considering the
belief and the actions based upon it irrational). Hence, any
action directed toward those (irrational) objectives or
resulting from a desire to follow those norms will also be
irrational.
If this is true, a great deal follows. Persons acting largely or entirely on the basis of emotions or feelings are often said to be acting irrationally. This fits in with what has just been said about irrational action, for such persons can be seen as acting on the basis of beliefs about what they want or about the situation they are in that are not supported by the evidence available to them. Emotions or feelings in and of themselves do not count as good evidence for a belief, and so an action based on a belief which is supported more by emotion than by the facts will be irrational. This needs to be qualified, however. Since all actions aim at some object, and those objects most primary and basic (in that all other objectives may be ways of achieving the primary objectives) consist in the fulfillment of some elementary desire (and so those objects are, as it were, simply a matter of fact for the agent, and not supported by other beliefs or evidence simply because they are in need of no such support), then there will be a component of feeling in all actions, at some level. Or so Hume argues, and he seems to be right. It seems that if one is aware of what the basic or primary objectives and values one has are so that these then become a matter of choice rather than fact, one has attained fuller agency and greater freedom (as the Existentialists point out)\(^{11}\), but the absense of this awareness is not in itself an aspect of irrationality (although it could be if the objective were actually counter to the agent's real and deeper interests and existed as a disguised perversion of those interests, as may be the case with masochism, for
example). There may even be cases of actions which are morally praiseworthy and which have been guided by a feeling, such as affection or compassion. (One should note, however, that in such cases the feeling itself is morally praiseworthy only insofar as it is founded on beliefs which do not run counter to the facts present to the agent.)

Yet there are cases where a feeling or emotion distorts or excludes the facts in such a way that the beliefs of the agent are made irrational by the presence of those feelings. In such cases, actions based on beliefs distorted by feelings would be irrational. (Emotions can distort belief either over a short period of time, as when one's view of the world is distorted by anger or sadness or infatuation, or the distortion can be chronic due to the constant presence of the emotion, as is the case with prejudice. The process is self-reinforcing, as the distorted beliefs provide a basis for the emotion, which continues to distort beliefs, and so on.) This conforms with the fact that in the absence of strong emotions, or after feelings have subsided, the agent is likely to see the situation in a different way and to act differently as a result.

Still, our account is incomplete. There are times when an agent may act on an emotion that is not founded on the evidence, but is not founded on the agent's belief either, and, moreover, does not affect the agent's beliefs. This occurs with persons suffering from phobias, who know they have no reason to fear the object which they fear. So irrationality is not simply a matter of a disparity between beliefs and evidence, but also may be a matter of disparity between evidence and
emotions, or even between beliefs and emotions. The discordance between attitude and evidence may produce irrational actions. Indeed, there may even be cases where the absence of certain emotions is irrational if certain beliefs are present, as may be the case with psychopaths and sociopaths. These cases may be analogous to those where an agent is lacking in a belief when all the evidence necessary to support that belief is available to the agent. The discordance can thus be either the presence or absence of an emotion, in the former case when the agent's beliefs do not support the emotion and in the latter case when the agent lacks an emotion his or her beliefs support. Actions based on such discordance will be irrational (but the matter is one of degree, with greater discordance providing better grounds for regarding the action as irrational and slight discordance as being perhaps insignificant).

Certain beliefs in extraordinary powers, such as telekinesis, would not necessarily be irrational, even if the basis for such beliefs may be largely emotional (as may or may not be the case) if such beliefs do not run counter to the facts available to the agent or distort other beliefs. In cases where feeling and belief are to a degree inseparable (as in some forms of magic, religion and other "ecstatic" experiences), those feelings or beliefs would be irrational only where their presence distorted other beliefs or where such feelings produced beliefs that are incompatible with the evidence available. One must be careful not to overextend the concept of rationality so that any belief or action not
in concord with a dominant world view is classified as irrational. Irrational beliefs are only those which, because of their falsity and intractibility, in some way or another incapacitates the agent (to a degree) in the agent's attempts to achieve his or her goals (by causing the agent to choose unsuccessful means, or to act against the agent's own best interests). To the degree to which a feeling or belief so incapacitates an agent, an action based on that feeling or belief will be irrational.

Rationality also involves the ability to consider the relevant evidence. If an agent acts in such a way that he or she achieves a secondary goal at the expense of a primary one, this action could either be irrational or simply mistaken. If it is pointed out to the agent how such an action defeats his or her primary goal and the agent then amends his or her actions in order to avoid this undesirable result, we would say that the agent was acting rationally but mistakenly. However, if the agent failed to amend the action in the light of the new evidence, to the degree to which the evidence is compelling and to which the agent resists adjusting the action we would say that the agent was acting irrationally.

This thesis places self-deceptive beliefs in the class of irrational beliefs, and so classes actions based upon self-deception as irrational. Self-deceptive beliefs run counter to the evidence available to the agent. One form of self-deception is deceiving oneself as to one's reasons for acting. One can be self-deceived about one's own motives and intentions. Therefore, in order for an action to be rational,
it is necessary that the agent be aware of the real reasons for the action, for among the beliefs of the agent which must be rational in order for the action produced by them to be rational are those which concern the agent's motivations. (Again, it is a matter of degree, with greater self-awareness leading in the direction of greater freedom and fuller agency, and greater discordance between one's real reasons and one's assumed reasons leading in the direction of irrationality.)

However, one can imagine cases where the agent is able to incorporate new evidence into his or her systems of beliefs consistently, where the agent's goals and the ordering of those goals are consistent and where the agent can give a true account of his or her reasons for acting which we would still call irrational. There could be a delusive paranoid (or at least, someone who appears to be one) who believes she is Joan of Arc reincarnate, and whose system of beliefs is such that any evidence will serve to reinforce that belief, given the interpretation she gives the evidence and which is supported by her system of beliefs. Perhaps we are driven to some kind of epistemological relativism here, but this need not be. Perhaps it is the fact that any evidence will support her belief that she is Saint Joan that makes such a woman irrational. That is, it is the rigidity of such a belief that makes it irrational, and that distorts all the other beliefs of that person. So it may be a necessary condition of rationality that a person be at least capable of altering certain basic beliefs if the evidence would be
better handled by such an alteration. Yet, there remains the problem of what it is to handle the evidence better. If all that means is that the best way to handle the evidence is that way which best achieves one's goals, then Saint Joan, whose goals are consistent, may well be able to achieve her goals best by believing she is Saint Joan. She may be able to achieve her real needs best that way, and not just her goals, so that she lives a long and happy life. It may be the case that she is in fact Saint Joan, but only she is in a position to know that (God talks to her and her alone about her identity). It would make no difference to us if she really were Saint Joan if only she could know that: we would still call her irrational. (Logical positivists might want to go so far as to assert that she was speaking nonsense.) But we would not be just in doing so, for unless we could show how her system of beliefs is inconsistent, or how she fails to achieve her goals (perhaps due to an inconsistency in the belief system), calling Saint Joan irrational is a form of epistemological tyranny: it is the imposition of our world view on another world view that is internally consistent and as able to deal with the facts as our own.

Although such a person is not irrational (and therefore not excusable on those grounds), such a person is indeed living "in another world," with another belief system and hence with another set of rules. (Even if the values were basically the same, the rules would be different as a radically different belief system would propose different means for realizing those
values.) To try to make such a person conform to our rules, which are based on a different belief system, would be foolish, useless and perhaps even unjust. Since responsibility is a practice based on our beliefs and rules concerning human action, it would not be possible to hold a person with a radically different belief system responsible, for such a person could not participate in the practice. In contrast, although irrational agents are not held fully responsible for their actions, such agents are liable to be treated by us to correct their irrationality and thus alter their behaviour because such agents have inconsistent belief systems, or are not able to handle the evidence sufficiently well, or cannot give a true account of why they acted, or continually fail to achieve their most important goals (and these factors are clearly related). By treating such agents so they will act more rationally, and hence with greater agency and responsibility, we enable those agents to better achieve their goals. A rational agent acting irrationally in a particular circumstance would not be held responsible for the irrational action, as an action based on irrational (and hence mistaken) belief does not reflect the agent's true desires and intentions, and would in a way be similar to a mistaken action (and similarly excusable). In the case of both the irrational agent and of the irrational action, the agent is subject to treatment in order that the agent may in future be able to act responsibly, that is, with standard agency. Such would not be the case with Saint Joan; she is not a participant
in the practice of responsibility, and any attempt to make her join the practice by conforming to our world view would be unjust, in that it would not be treatment, as she is able to realize her goals, but coercion. Rationality is then another component of standard agency, and hence a precondition for full responsibility. Participation in the same basic belief system upon which the practice of responsibility is based is a necessary precondition of any participation in the practice of responsibility at all.

6. Involuntary actions

We have looked at ways in which intentional acts may go wrong and at ways in which intentions themselves can go wrong in such a way as to diminish or eliminate responsibility for an action. The next question a theory of action must address is whether persons can act intentionally but "against their will." Can a person perform an act that is both intentional and involuntary?

The prime example of acting "against one's will" is coercion. In coercion, it is argued, one is forced to accept a goal or an ordering of goals that would otherwise not be part of the interrelated set of needs and desires of the agent. In other words, circumstances make one accept a goal one would not otherwise have held.

There is a trivial way of construing this formulation, and a trivial objection. Circumstances always act as a determinant of the goals of the agent, even in cases where the
action is not coerced. If I knew no one in Saskatoon, I would not visit there. However, as I have relatives there (whom I want to visit), circumstances are such that I adopt a goal that otherwise I would not hold, i.e., the goal of visiting Saskatoon. Clearly, this is not a case of coercion or involuntary action, and so the factor which marks coercion off from other determinants must be something other than the one cited in the formulation given.

There is, however, another way of approaching the problem. It could be argued that voluntariness is having standard conditions of choice. Involuntariness is a deviation from standard choice and occurs under conditions that deviate from the standard. Not being free is being subject to overriding conditions which are not normally in force. The overriding conditions (such as a gun to the head) produce an overriding choice (survival) which is a choice that it would be irrational not to choose. Normal conditions are not fully specifiable; they are the absence of abnormal conditions, which are specifiable. So voluntariness refers to choice, which is pre-intentional, and thus an involuntary choice (and an involuntary action produced by such a choice) is one not determined by normal conditions in which the agent's normal goals and the ordering of those goals may come forward as determinants of the agent's action. An involuntary action could then be produced by coercion, by extraordinary circumstances, or by an organic malfunction such that the agent's knowledge and preferences do not determine the intention or
the action.

There is something strange about this line of argument, though. Take the examples of coercion and extraordinary circumstances. (The former seems to be a subset of the latter.) In such cases, the agent's goals and goal ordering, far from not determining the actions, become clearer than normal and determine the action to a more obvious degree. The agent's lower order preferences do not come forward as determinants of the action precisely because his or her higher order preferences do. Take for example being told that one must eat pork or have one's brains blown out (the latter certain to result in death except for those who enter politics, as Mark Twaine pointed out), and the situation is such that those are the only alternatives (one cannot overpower the person with the gun, or escape by some ruse). If one dislikes pork, and even if one has a strong moral prohibition against eating pork (as would be the case with members of some religious groups), one is more likely to choose eating the pork than to choose death. But the point is that, whichever of the two choices one makes, one makes a choice, and so acts on the basis of one's preferences. Certainly the situation is such that one acts on one value (self-preservation) at the expense of another (abomination of pork eating, or proper reverence to religious authority or some supreme being), but one chooses which value has priority, and so by acting on a priority of values one has not acted "against one's will" at all. It is therefore not the case that involuntary action is not due to the system of causes
and potential causes (such as preferences and goals) which defines the agent, for it is precisely under circumstances of extreme stress (where one is forced to choose between highly-ranked values) that one may discover just which values one really holds highest. There is no such thing as an involuntary choice when the notion (of involuntariness) is used in this sense, as one never chooses against one's will, and so there is no such thing as an action which is both intentional (that is, the product of the agent's choice) and involuntary. At best, the notion of an involuntary action can refer to actions produced by extraordinary circumstances where very highly valued goals determine the outcome of the action, but even so, the word "involuntary" is misleading.\textsuperscript{14} In the vernacular, involuntary emphasizes the presence of extraordinary circumstances, but it is a mistake to think a person who acts intentionally and on the basis of choice under such circumstances as acting "against their will." In extraordinary circumstances one still has a choice, and there may be cases where one very highly valued goal (eg., self-preservation) is sacrificed to another even more highly valued goal (eg., love of another, love of country, strong feelings about murder, etc.). The agent still chooses, and the choice is a product of the agent's character (which is nothing more perhaps than a disposition to behave a certain way in certain circumstances). Any intentional act which is a matter of rational choice is then voluntary.

There can be no grounds for excuse for an intentional action simply on the basis of the presence of extraordinary
circumstances or a situation of coercion, for nothing has gone wrong with the agent's action (assuming the intended result was achieved), nor (we assume) did the agent act irrationally. This is a surprising conclusion considering that persons acting under coercion or extraordinary circumstances are very rarely punished. But the surprise fades when one sees that the reason they are not punished is not because such persons are not held responsible. Indeed, in such cases the action is not excused so much as it is, in a way, justified or condoned, for the action took place under circumstances where the agent could have avoided violating one value (eg., the prohibition against pork eating) only by violating a stronger or higher one (eg., self-preservation), where the stronger value is one which we could not reasonably expect the agent to violate (simply because to place such a high demand on human beings would be fruitless, as the demands would not and perhaps could not be met most of the time, and so a rule enforcing such demands would be inefficacious.)

To perform a coerced action then is to perform an act that the non-performance of that act would have caused one to violate a goal or value which one could not reasonably be expected to violate. The values in question are usually very basic and rational desires (such as self-preservation). Thus, when a person acts under coercion, we do not so much excuse the action as admit it was morally permissible (and in some cases morally right) under the circumstances. (There may be cases where the agent's value structure itself cannot morally justify
the action as that value system may not coincide with the value system of morality, but coercive interference mitigates culpability as it presents the person with a choice he or she need not otherwise make. The presence of coercion may make the person perform the morally wrong choice, but the choice cannot be condemned, at least not very strongly, because the consequence of the morally right choice is the violation of a value so high in the agent's priorities that we could not reasonably expect the agent to violate it.)

Another sort of case subsumed under the concept of involuntary action is that of compulsion. It is clear from our earlier analysis of rationality that this kind of action is better included in the category of irrational actions, for compulsive behaviour (whether induced by organic malfunction, such as a brain tumour, or by a purely psychological disorder) is behaviour produced by a system of causes within the agent that is malfunctioning. The compulsive really has no choice, whereas the person who is coerced does. It is part of the definition of a compulsive that the agent cannot choose to not act on the compulsion. In a sense then, compulsive actions are involuntary, as the agent has no choice. This may be due to an inability to consider the evidence, or to an overriding emotional factor, or some other cause that produces incoherence in the agent's belief system, or in the relation between the agent's knowledge and the agent's emotions (where the emotion may produce behaviour). The compulsive does not choose against his or her own will, though, for it
is incorrect to say that the compulsive chooses at all. As with other kinds of irrationality, the matter is one of degree. The stronger the compulsion, the narrower the range of choice or the lesser the ability to choose, but there may be some compulsives for whom choice is not non-existent but is merely severely limited by the strength of the compulsion.

There are two types of involuntary actions which remain. The first involves persons being physically forced to do something, as when someone is pushed. The second involves those movements such as muscular spasms and twitches which are said to be involuntary. While in both types of cases the movement of the person is in a sense involuntary, in neither does the movement count as an action of that person.

Take as an example of the first sort of case a person being pushed into a glass door so that the glass breaks. One might ask who broke the glass. On the face of it, the person pushed into the door broke the glass. Analogously, if a rock is thrown through a window, one can say that the rock broke the glass. But the rock did not act. It was the action (throwing the rock) that caused the glass to break. So it is in the case of the person being pushed into the glass door. That person performed no action. While in a sense that person broke the glass (as in the sense the rock breaks the glass) that person did nothing to break it and therefore would not be held responsible for the breakage. The breaking of the glass would be an action of the person doing the pushing described with reference to its effects. The event of the
glass breaking was not produced by a system of causes within the person being pushed, but was produced by the action of the person who pushed that person. Thus, it was the person who pushed who broke the glass, by pushing someone else. The person pushed is not responsible because that person does not act.

A person physically forced into movement does not act, then, but is acted upon. The movement is not produced by a system of causes within the person, but is either a simple event (produced by non-human forces) or is the effect of the action of another (and is in that case produced by a system of causes internal to an agent, but not internal to the person who is moved by physical force). The person moved is not responsible for the effects of the movement, as the movement was not an action of that person. A person physically forced into movement moves involuntarily, but does not, properly speaking, act.

One can imagine circumstances where a person physically forced into movement might be held responsible for the effects of that movement, as when a person should be on the alert for physical events and human actions that might force that person to move against that person's will. But the person in such instances is held responsible for being in a situation or for not being alert, and not for the actual movement. While the movement produces the effect for which the person is held responsible, it was the person's presence in the situation or the person's lack of alertness which allowed the movement to
take place. The person would be held responsible for the effects of the movement because the person was negligent in that the person had control over being in the situation or being alert, even if the person had no control over the actual movement. In effect, the person is held responsible for the effects of negligence, and not, strictly speaking, the movement. The movement would not be an action, and so the person would not be responsible for its effects in the way he or she would be for an action (which is produced by a system of causes within the person). The person would only be responsible for the negligence.

Physically forced movements differ from unintentional actions in that the latter involves intentions not being realized (and so not being causally effective) while in the former there are no relevant intentions operative at all. It is not the case with forced movement that an action miscarried so that an intention was not realized (as is the case with unintentional actions), for the movement itself was not intended and not an action. There was no intention present to be realized. It is not a case of an action breaking down, simply because it is not a case of action.

Involuntary movements such as twitches and spasms at first seem to differ from physically forced movements in that they are produced by a system of causes within the agent. But they are not actions because no intentions are operative, and so the movement, and not simply its effect, is unintentional. A movement, we have seen, becomes an action when it becomes subject to or results from the agent's deliberate and conscious
control. In other words, there must be an intention operative for an event to be an action. While twitches and spasms are produced by a system of causes within the person moving, the system of causes is not the sort necessary for the movement to be an action because it lacks intentions. One is not responsible for twitches and spasms as they are not actions. One may be responsible for the effects of such involuntary movements in certain situations -- but as is the case with physically forced movements, one is then only responsible for being in a situation and for one's awareness, not for the movement itself. One has control over the former, but not the latter. Twitches and spasms are involuntary in that the person experiencing them has no control over them, and so they may occur against the person's will, but they are not actions.

There is a confusion between involuntary movements and coercion, for in both sorts of cases we may say that the person "couldn't help it," but this is in fact only true of involuntary movements, because they are not a product of the person's choices and intentions, whereas coerced actions are.

7. Conclusion

We are now in a better position to understand what an action is. It must be produced by a system of causes with the agent. This system of causes must include some intention of the agent, as the difference between an event that takes place in the body or involves the body and an action is that
the latter is the product of the conscious exercise of power.
These are minimal conditions for an event's being an action.
It is what remains of action after whatever adverbs of excuse
might indicate that the intentions of the agent have not been
realized, and thus that the action broke down at some point.
The intention that produces the action is not realized when
greater causal factors determine the outcome of the action.
The realization of intentions is a necessary condition of
successful action, or action in its standard, unqualified
sense. It is thus a necessary condition of standard agency
and responsibility. But it is not a sufficient condition,
as full agency requires that the agent be rational, i.e., that
the agent's goals, beliefs, and emotions be coherent. Full
responsibility then requires that the intentions of the agent
be realized and that the agent be rational.

There is a pragmatic basis for the concept of responsi-
bility in the concept of agency. Persons are held respon-
sible for those things over which they have the power to
control. When an action breaks down or fails, the agent lacks
control over its results or effects and so is not responsible
for it. Moreover, the agent is not held responsible for it
to the degree to which the agent had no power to control it.
The agent may be held responsible for those things that would
have ensured the realization of the agent's intentions, such
as proper foresight, attention to the situation and to the
manner of performance of the action, when these factors are
subject to the agent's control. In pointing out how an action
breaks down, one reveals to the agent the steps that can be taken (or should have been) to ensure the realization of intentions. In effect, one makes the agent more fully aware of the areas over which the agent may (and is expected to) exercise control, thus facilitating the agent's control over those areas. One thereby increases the freedom of the agent and enlarges the area for which the agent is responsible.

An irrational agent is not held responsible because the actions and intentions of that agent are not subject to that agent's control. In treating an irrational person, one seeks to increase the person's agency by developing in that person the capacity for rational choice and action which brings the intentions and actions of that person under the person's deliberate control. In doing so, one gives the person a capacity for agency and hence for responsibility.

The simple reason that responsibility rests on agency is that it is a practice designed to control behaviour. It does so by pointing out to agents their agency. Responsibility can only control behaviour if the behaviour is subject to the control of the agent. It is only common sense to ask agents to control only that which they can control. Responsibility (as a practice) is an attempt to control behaviour as it asks the agent to recognize not only the good or bad consequences of an action, but the agent's power to produce or prevent such consequences. In holding someone responsible one is claiming or treating the outcome of the actions as one over which the agent had control. Responsibility is thus a recognition of agency and a demand addressed to the person being held
responsible that the person recognize his or her agency. One makes actions subject to moral judgments and extends those judgments (and the measures reflecting them) to the agent by pointing out to the agent that the moral quality of the action was a factor over which the agent had control. One thus demands that actions be subject to moral standards and sanctions, thus controlling behaviour. The more fully agency is developed in this way, the greater is the responsibility of the agent. Hence, the greater is the moral regulation of the agent's behaviour.

The practice of responsibility has pragmatic ends other than the control of behaviour (which could be accomplished by other means, such as shock therapy or other behaviour modification techniques). The difference between responsibility and other practices for controlling behaviour is that responsibility, in recognizing agency, recognizes the control of an agent over the result of an action, which is a recognition of the agent's freedom. In developing and pointing out agency, the practice of responsibility indicates ways in which we may act with fuller control over our actions, that is, more freely. To have agency is to have control, and to have control is to be free. Our analysis of agency has thus given us not only an understanding of responsibility, but of the notion of human freedom as well. That freedom is a matter of control over one's actions and that it exists is something for which we must now argue.
1. Introduction

There is a view which holds that there is something fundamentally wrong with the account of responsibility given so far. According to this view, no one is ever responsible because no one has the freedom necessary for responsibility.

The argument runs as follows. One is responsible only for those actions which are free; if one could not help doing as one did, one is not responsible for the action. Thus, if one is responsible for an action, it must have been possible for one to have acted otherwise than one in fact did. If one could have acted otherwise than in fact one did, then one must have had the power to have realized a different choice or intention than the one acted upon. Furthermore, one must have been able to have chosen or intended otherwise in that situation than one did. However, the argument continues, one is never really able to act otherwise than one does. One's powers, capacities and one's proclivities for choice are causally determined by one's genetic make-up, environment and previous experiences, all factors over which one has no choice or control. Since how one acts is a product of one's powers and choices, and those powers and choices are causally
determined, how one acts is causally determined. Therefore, concludes the argument, one cannot act otherwise than one in fact does, and thus one is not free in the way that is required by responsibility. Hence, it is said, no one is ever responsible.

While this argument does place limits on the concept of responsibility, particularly on the practices of praise and blame, it does not rule out the concept of responsibility that has been argued for here.

Determinism does not rule out freedom because the sense in which it must have been possible for an agent to have acted otherwise than the agent in fact did in order for the agent to be free and responsible is not that given the exact same causal antecedency the action could have been otherwise, but whether the causes of the action were such that the action was in the agent's control, so that if the causal antecedency had been different in some respect than did not change the fact that it was the same agent acting in a relevantly similar situation, the agent had certain powers and abilities that could have been exercised differently. Given the exact same causal antecedency, no event could have been different than it was, and so if an agent could have acted differently, something in the causal antecedency must have been different. This is analytic, and so does not depend on any principle of universal causality, as even if the action was caused by the agent's free will, that free will would have had to have decided to act in the same way if the causal antecedency of the action is to be exactly the same. It will be argued, therefore,
that the postulation of an uncaused or free will does not explain freedom of action, and that the concept of freedom must be understood in terms of the individual's control over the action, which is a matter of the powers and abilities of the individual, and so of what the causes of an action are, and not of whether or not the action was caused.

2. Defining the problem

It will be useful to begin by considering some previous attempts to deal with this argument. While none of them has been very satisfactory, they do indicate what the proper approach to the problem should be.

One argument, put forward by P.H. Nowell Smith, claims that to say that an agent could have acted otherwise is to express a hypothetical proposition indicating that if the agent had been slightly different, or if circumstances had been different, then the agent would have performed a different action. If this analysis is correct, then certainly the freedom needed for responsibility exists, as the hypothetical proposition might often be true. There would then be no problem about responsibility. Similar arguments have been made that state that to say the agent could have acted otherwise is to express the hypothetical that if the agent had wanted or intended to act otherwise, then the agent would have done so.

Unfortunately, Nowell Smith's analysis is hopelessly flawed, as is any other attempt to put the requirement that the agent could have acted otherwise into a hypothetical form-
C. A. Campbell, in "Is Freedom a Pseudo-Problem?"\textsuperscript{16} rightly notes that such analyses do not meet the point at issue. Campbell points out that in assessing whether an agent is responsible, we assess whether the action of that particular agent, in that particular situation, was free. We are concerned with assessing the freedom of that particular agent, and so are concerned with whether that agent could have acted otherwise in the same circumstances. We are not interested in what a different character would have done in the same situation, or in what the same character would have done in a different situation.\textsuperscript{17}

Campbell, however, is unclear as to what is to count as the same character or the same situation. There is some truth in the intuition behind the hypothetical analysis of "could have done otherwise." That is, when assessing whether an agent could have done otherwise, one is assessing the agent's capacities to have done otherwise, and an assessment of capacities may be expressed hypothetically, although it would be misleading to formulate the assessment in that way. To ask if an agent could have performed an action is to ask whether it would have been true to have said in that situation that "the agent can perform that action." The question is then essentially whether the agent has (or had) the capacity to perform a certain action. It is not to ask whether the agent would have performed the action given exactly the same conditions, for that is not to ask whether the agent has a certain capacity, but whether, given certain circumstances,
that capacity would be actualized. Again, compare the question of whether it would have been true to say in a given situation, "the agent will perform that action," with the question concerning the truth of "the agent can perform that action."

The questions are distinct, even though, as the intuition behind the hypothetical analysis recognizes, a definition of capacities may be to a certain extent be expressed by specifying what a thing or person may do under circumstances. Yet, it is misleading to express this in terms of what a thing or person would do, for it is not a question of what a thing or person would do, but of what a thing or person is capable of doing. For example, to say that a car could have gone 50 kph is to say that it was capable of doing so, not that it would have done so given the exact same circumstances, which may include the fact that the accelerator was far enough down that the car went twice that fast. It would be incorrect to say that the car was incapable of going 50 at that moment; in fact, it went faster, but it could have gone slower. Which is not simply to say that the car would have gone slower if the accelerator was not pressed so far down, although that hypothetical would usually be true. It does indicate the capacities of the car by claiming that the same car in the same condition on a road of the same condition and difficulty could have gone 50. Obviously, a damaged car could not have gone 50, and a bad, treacherous or blocked roadway would have prevented an undamaged car from doing so. Thus, the capacity is defined by specifying the condition
which must be true of a thing or person for the capacity to be present, and by specifying the condition which must be present in the situation for it to be possible for that capacity to be actualized, one specifies what constitutes an opportunity for the exercise of the capacity. To indicate what could transpire when certain conditions are fulfilled is to indicate what the capacity is; fulfillment of the conditions constitutes the truth of the statement that the thing or person could have done the action made possible by the presence of the capacity in question. To go beyond those necessary conditions is to specify which conditions would lead to an actualization of the capacity; and fulfillment of those conditions constitutes the truth of a statement about what a thing or person would have done. Though actualizations of a capacity make the presence of a capacity indubitable, the absence of an actualization, where the necessary conditions for the actualization of the capacity are fulfilled, does not indicate the absence of the capacity. Only an absence of actualization in the presence of conditions that should be sufficient for its actualization would indicate the lack of the capacity (although it would take repeated tests to determine that the capacity was in fact lacking, and that its non-actualization was not due to some other factor, such as a temporary malfunction or the overriding of the system of causes within the agent by some greater causal factor that prevented the agent from exercising the capacity). If the accelerator was depressed and the car did not respond, it lacked a certain
capacity (assuming the engine was on and the car was unimpeded by obstacles). If a desire which is strong enough to motivate an action is present but either cannot be translated into action, or the action is unsuccessful despite the presence of conditions that would make a successful action possible, then the person who has that desire either lacks a capacity or that capacity is limited (depending on how one wants to describe the capacity, since absence of a capacity may sometimes be described as presence of a similar but lesser capacity). Hence, to say that a person could have done something is to say the person had the capacity to do it, not that under the same circumstances the person would have done it. It is to say that a person was able to have performed an action, given that the circumstances presented the opportunity for the performance of that action. 18

3. Capacities and freedom

We can now apply these considerations to Campbell's objection to formulating the statement that the agent could have done something in terms of what the agent would have done. Although in assessing the responsibility of an agent one is not interested in what another agent might have done in the same situation, one may be interested in whether, in a situation presenting the same opportunities for action, an agent could have acted differently if a condition that is usually sufficient for a different action which the agent has the capacity for had been present. (For example, a strong
desire would be a usually sufficient condition for an action in that it would produce an action in any agent having the capacity for the action.) Lack of action in such circumstances would strongly indicate the lack of a capacity, in which case the agent would not have been free to have performed that action. Thus, one may be interested in what factors functioned as motives in producing the action and in whether the agent had the capacity to act or to refrain from acting on those motives. One is thus interested in whether the conditions which must be true of an agent for certain actions to take place are fulfilled. One may be interested, for example, in what mental factors are necessary conditions of actions, and in whether an agent is capable of fulfilling those conditions. One might then be interested in how the same agent in a different frame of mind might have acted. It is relevant to an assessment of the capacities of an agent to consider whether the agent would have acted differently if the agent had been more alert (which is to ask whether the agent had the capacity to be more alert, and thus to ask whether the agent was somehow negligent), or if the agent had been more reflective and less emotional (which is to ask whether the agent acted rationally or was capable of doing so). Certainly, this is a consideration of the same agent and not another. In assessing whether an agent has certain capabilities, one considers not whether the agent would have acted differently given the exact same causal antecedency of the action (for given that antecedency, the agent could not
have acted otherwise and therefore had no capacities to do so), but whether, if the causal antecedency had been different in some respect, such as the agent's frame of mind (but not in those aspects that constituted the opportunities present in the situation), the agent had certain powers and abilities that could have been exercised differently in that situation.

If someone has the capacity to perform an action, then that person also has the capacity to not perform it. Otherwise, what occurs is not the result of a capacity for action. (So, for example, to the extent to which I am incapable of not breathing, my breathing is not an action, but a reflex.) Actions must include in their causes an intention, and a movement a person is incapable of not making is therefore not an action. An action one is incapable of not intending is an action, albeit a compulsive one, and an agent who acts compulsively is incapable of doing otherwise, in which case it is not that the agent is capable of the compulsive action (though in the sense that the agent is capable of realizing the compulsive choice the agent is capable of the action) so much as the agent lacks the capacity to not perform that action (as the agent lacks the capacity to not choose that action). To ask whether an agent could have acted otherwise is to ask whether the agent had a capacity for doing so, not whether that capacity would have been actualized (as it clearly was not) under the exact same circumstances.

This analysis of what it is for it to be true that an agent could have acted otherwise thus indicates that freedom is
a matter of one's capacities to act. In other words, it is a matter of agency. There are two sorts of capacities which must be considered: the capacity for action and the capacity for choice. The capacity for action is the freedom or ability to translate choice into action. It is necessary for responsibility, for without it one's intentions could not be realized. The capacity for choice is the capacity to control one's intentions by making them subject to one's interests on the basis of one's knowledge. It is thus what we have referred to as rationality, or the capacity to be rational. (One can choose irrationally, in that one can intend irrational actions and so act irrationally, but as such choices are not under the control of the agent and so not free they do not result from a capacity for choice, as to be capable of choosing something is to be capable of not choosing it, and one cannot be capable of not choosing an irrational choice, since one is not capable of not intending irrational actions, as those actions are not subject to one's conscious control. By the capacity for choice we thus are referring to the capacity for free choice, or choice that is under the agent's conscious control and thus rational.) One could have the capacity to have realized a different choice even if one's choice and hence one's action was causally determined. Similarly, one could have the capacity for choice, in the sense that one's choices were subject to one's rational direction or control, even if the choice one made was causally determined. To be able to have acted otherwise is to have the capacity for
choice and action, and that is the only freedom necessary for responsibility. Actions which are unintentional are not subject to the agent's control, and so the agent is relieved of responsibility for them. Irrational actions are those produced by intentions over which the agent has no control (and which may in fact run counter to the agent's best interests) and so are not subject to the agent's control and thus are not actions for which the agent is responsible. The question is then whether an agent is really ever in control of an action, the answer being that the agent is in control when the agent has the capacity to choose and act otherwise.

4. Incompatibilist objections: Campbell and Hospers

But those who argue that determinism and freedom are incompatible are not going to give up yet. They argue that if determinism is true, then ultimately no actions are subject to the control of the agent, since ultimately they are caused by factors beyond the agent's control. Campbell, for one, argues that if one's choices and powers of agency are causally determined, then any action resulting from causal determination is not really in the agent's control, since the causal factors which produce the action (via the agent) are not within the agent's control.

Before assessing the merit of this argument, one might look at the solution Campbell proposes. The way out of what Campbell takes to be a dilemma is to assert freedom and deny universal causation or determinism. Campbell urges us to suppose we have a self which has a contra-causal sort of
freedom. This kind of freedom is necessary for responsibility if one accepts Campbell's argument about determined actions. According to this account, the Campbellian self is responsible for that "effort of will" experienced whenever we do what we believe to be moral, rather than what we desire. A charming notion, surely, but a relic of the dark days of British metaphysics. It is, in the first place, entirely arbitrary. Campbell provides no convincing rationale for separating the desire to be moral from other desires, but this separation is the basis for his separation of the uncaused "self" from the causally determined "character," the former opposing its "effort of will" to the latter's desire. Without such a separation there remains no effort of will opposed to desires and no "self" distinct from "character." The notion of an uncaused self moreover makes no sense; it flies in the face of psychology and common sense. Whether an individual makes an effort or tries to do something is a result of that person's causally determined character. For example, the person's sense of duty, desire to achieve and so on, which produce the "effort of will" we experience, are products of experience and environment, such as parents, teachers and other figures of authority interacting with that individual. Whatever the "effort of will" is directed toward as a goal is also a result of aspects of one's character that are causally determined. The "effort of will" must have an end, and that end is the result of causally determined choices and preferences. An "effort of will" without purpose is senseless, and
a purposive "effort of will" without causes is equally senseless. Moreover, if freedom exists only in opposition to desires, then one is most free when one resists all one's desires. Yet surely such a condition is pathological, and would indicate the absence of freedom (since it indicates some sort of compulsion), rather than the presence of freedom.

Campbell thus leaves the problem he poses unsolved. The problem is that if the causal factors which determine an action are beyond the agent's control, then the action is not within the agent's control and therefore the agent is not responsible for it. The extent to which the antecedent is true and whether the consequent follows from it is something which needs examination.

A clear and forceful statement of the problem can be found in Hospers' "What Means This Freedom?". Hospers begins by pointing out that "frequently persons we think responsible are not properly to be called so." Many actions may result from unconscious drives over which the person has no control, and while they appear deliberate and intentional, they cannot be affected by "reasoning, exhorting or threatening." They are not subject to the agent's conscious control. Indeed, the behaviour which Hosper's describes is that which we have labelled "irrational," and it has been shown that the agent is not responsible for it. But from the fact that agents are not responsible for their irrational behaviour, Hospers falsely concludes that agents are not responsible for irrational behaviour because they are not responsible for the situations
of childhood which produced that behaviour. 21 In fact, the agent is not responsible for such behaviour simply because it is not subject to the agent's conscious control. Why it is not subject to the agent's conscious control may be relevant to considerations of how to go about making it subject to the agent's conscious control, thereby increasing the agent's responsibility and agency, but it is not relevant to considerations of whether or not the agent is responsible. (In other words, it is a purely therapeutic concern based on the presumption that the agent was not free and not relevant to the question of whether the agent acted freely.)

Consider Hospers' argument: One is not responsible for one's childhood situations. One's childhood situations produce neurotic or irrational behaviour. Therefore, one is not responsible for one's neurotic behaviour.

Now consider a parallel argument: One is not responsible for one's childhood situations. One's childhood situations result in rational behaviour. Therefore, one is not responsible for one's rational behaviour.

There is as little need to accept the second argument as the first, which is to say none at all. One is not responsible for irrational behaviour because it is irrational, not because one is not responsible for the childhood situations that resulted in one behaving irrationally. Responsibility is a matter of agency, that is, a matter of control. The rational person's actions are subject to that person's conscious control, and therefore that person is responsible for them.
(The concept of rationality is that outlined in Part One, which is basically the coherence of goals, beliefs, emotions and the information available to the agent.) The irrational person lacks such control, and so lacks responsibility. Even looking at childhood situations for which we are not responsible, we are not responsible for them because we lack power and control in them (and to the degree to which we lack such control), not because they are caused by factors beyond our control. This indicates that responsibility is not a matter of control over the causes of an action, but of control over the action. It is one thing to say that agents have no control over the factors that created their characters, and hence determined how they behave, and it is another to say that those agents have no control over their actions.

Hospers extends his account to exclude all persons from responsibility, and not just those who behave irrationally. That such a move is illegitimate is a consequence of Hospers' wrong account of why we excuse irrational behaviour. From the fact that we excuse behaviour when we learn that it is produced by psychological conflicts, or that in some way irrational, Hospers reasons that "the more thoroughly and in detail we know the causal factors leading a person to behave as he does, the more we tend to exempt him from responsibility." It is the inevitability of the behaviour, given its causes, says Hospers, which causes us to excuse the agent.

Such reasoning is simply false. We do not excuse an action to the extent to which we know its causes; it is not always the case that "tout comprendre, c'est tout pardonner,"
and in fact the reverse may sometimes be true (as when the real, dastardly and previously hidden reasons for an action are discovered). Knowing the causes of an agent's behaviour, far from excusing the agent, may in some cases lead one to ascribe responsibility to that agent if one discovers that the action was rational (in the sense we are using that concept) and intentional. It is not an action's being caused that excuses an agent, but the nature of those causes. If in examining the causes of an action we find the agent acted irrationally or unintentionally, then we excuse the agent. Discovering the causes of an action exempts the agent from responsibility only insofar as those causes indicate that the action was either unintentional or irrational, and so in some way not subject to the agent's control. In other words, when the system of causes within the agent that constitutes standard agency breaks down, is faulty or is over­ridden, the agent lacks control over the action and so is exempted from responsibility. The question then is not whether an action was caused, but what the causes were. More precisely, the question is whether the action was produced by a system of causes within the agent that included a rational intention which was realized (and so causally effective). If so, then the agent was in control of the action and is responsible for it. If not, then the agent, for any number of reasons, was not in control of the action, and so is excusable (to the extent that the action was not in his or her control). Criminals who are excused are excused because of irrationality, not because of determinism. (Those who can
be shown to be rational and whose action was the realization of an intention hence are not excused. Whether one can be rational and also a criminal, except through accident or misfortune, is a psychological question to which this thesis presupposes no answer. Clearly, however, one can be responsible for an action which, judged by some norm not necessarily shared by the agent, may be judged as bad without the agent being irrational.) If an action which was determined were rational, we would not take its causes for excuses.

We can now assess the argument that if actions follow from one's character and one's character is determined by causal factors beyond one's control, one is not responsible for one's actions. The flaw in the argument is the failure to see that the presence of only certain causal factors, and not causal factors in general, is a ground for excusing an action. Persons who are excused because their actions result from their characters are excused not because their character has been causally determined, but because it has been determined in such a way as to make the formulation and realization of their intentions in a rational (i.e., non-self-defeating) manner impossible. The argument that determinism rules out responsibility is false.

There is, throughout Campbell's and Hospers' arguments, some notion of the agent being forced into an action or being impeded in action by forces or obstacles which somehow stand in opposition to the agent. Our previous account of action (Part One) shows that some outside forces or states of affairs
may make certain actions impossible by placing certain physical constraints on what courses of action are available to the agent, or by taking away the agent's power to act. If the agent lacks agency, the agent is not free and does not act. As long as the agent is able to act, however, the act is a result of the choices of the agent. Though the situation places limits on the agent's choices, it does not diminish the agent's ability to choose or to act on a choice, and so does not make any actions which do take place any less free. The quality of the action will be assessed considering the choices available to the agent, but the agent is not excusable simply because the range of choices was not infinite or even optimum at the time of the action. (But considerations of the choices available may lead to actions receiving a different valuation than they would receive in a situation where the range of choices was different or wider. The more extraordinary the circumstances, the less we are entitled to expect or demand the sort of actions that would take place in ordinary circumstances. Yet one is accountable for actions that are performed under extraordinary conditions; sometimes actions are permissable under those circumstances which otherwise would not be. This gives the appearance of the agent being excused for those actions, when the action is not really excused but simply permitted (in that it is not condemned) under such circumstances, for the simple reason that a prohibition of it under such circumstances would place too great a demand on agents and so would be inefficacious and unreasonable, as we saw in Part One, in the section on involuntary
action.)

The idea of compulsion or constraint due to inner forces or states of affairs is similarly confused in Campbell's and Hospers' accounts. One is constrained or compelled by such forces when one is not acting rationally or when one is incapable of acting rationally, in which case one's choices are not conscious (as they are produced by some unconscious drive or altered by the presence of a factor such as drugs, strong emotion or pain) and one's intentions are not subject to one's control. In such cases agency and freedom are diminished, as has been shown, as there the range of choices is limited or even eliminated (in the case of compulsive behaviour), because the ability to choose is limited. Lack of control of intentions, which is evidenced in irrational actions, may be the result of inner forces, and hence constitutes a kind of inner compulsion or constraint due to inner forces. But a desire or preference does not diminish one's range of choices or ability to choose; it merely allows a choice to be made. If a desire or preference is taken to be an obstacle, that merely reflects the presence of another desire, or another choice, realization of which is incompatible with the desire or preference seen as an obstacle. How the conflict is resolved is also a matter of choice, specifically, choosing how to order and value certain goals. (If, at bottom, any ordering is based on a choice that is not based on any other choice, this is a logical necessity, and has nothing to do with determinism. This issue will be explored in Part Three.) To say that a choice of the agent somehow compels or constrains
the agent is to put the will of the agent in conflict with itself. A choice does not in itself diminish one's ability to choose. (If a choice is compulsive, that indicates that the ability to choose freely is impaired, but the choice is a result and not a cause of the impairment.) To be forced to do what one wants to do by the fact that one wants to do it is not to be forced at all. Hence, to be able to choose, which is what it is to be rational, and to be able to act on one's choices are the only necessary conditions of responsibility and freedom, for it is these which constitute agency.

There is a final point which Hospers makes, though, which is worth considering. Whether one is rational or not is a matter of luck. If one can consciously control one's behaviour, so that one may alter it on the basis of rational considerations, one is just lucky and can take no credit for that happy state of affairs. A good early environment is a matter of luck, but so is the ability to overcome an adverse early environment and the ability to respond to treatment, as they result from one's genetic makeup and early experiences. Even whether one is able to try harder to self-overcome is causally determined by factors beyond one's control, and thus is a matter of luck. This does not, as Hospers thinks it does, rule out responsibility, for one may be responsible as a matter of luck and still be responsible. If one can consciously control one's behaviour one is in a sense lucky, but one is also responsible. But the point does have bearing on the issue of praise and blame and the ascription of moral worth to self and to others.
For even if one is responsible for one's actions, one's preferences and values are ultimately the result of causal factors beyond one's control, and so it is a matter of good and bad luck whether those preferences and values are compatible with moral or with the standards of morality of a society or whether one can choose or learn to subordinate certain preferences and values to the morals of a society in order to participate in that society. Whether one's behaviour will earn praise or blame then is, in a sense, a matter of luck, even when one is responsible for such behaviour.

5. Praise and blame and responsibility

There is a close connection, remarked on by many philosophers, between a person's being responsible for an action and a person's being subject to praise or blame for that action. To praise or blame is to measure the quality of an action against a standard of how we expect, desire or demand persons to behave (depending on the importance of the kind of action). Holding a person responsible is thus linked to an assessment of that person's behaviour, and so to an assessment of that person's character. From our earlier analysis of agency, we can provide a rationale for such a connection. An agent is responsible for an action when the action is a realization of the agent's intentions (provided that the agent is rational at the time of the action). Any evaluation of an action for which the agent is responsible is therefore an evaluation of the agent's intentions, and the agent's intentions are a
reflection of the agent's character. If the agent acted irrationally, the intentions would point either to the agent's chronic irrationality or to a temporary irrationality due to circumstances (in which case the agent is said to have acted "out of character"). If the agent acted rationally, those intentions would be a reflection of the desires, beliefs and attitudes of the agent. An evaluation of an agent's actions is in those cases an evaluation of the agent's character. An evaluation of the agent's character on the basis of the agent's action would be inappropriate when the agent acts irrationally, for one of two reasons. Either the agent's intentions were not subject to the agent's conscious control due to special circumstances, as when the agent acts "out of character," in which case the action would not be truly reflective of the agent's desires and attitudes and so a character assessment based on that action would be false or inaccurate; or the agent chronically lacks control over intentions or actions, as is true of chronically irrational agents, in which case it would be inappropriate to place the same expectations on that agent as one would place on a rational agent, and so the standards against which actions are measured in praising and blaming would be inapplicable. (One might welcome or rue such actions, but to hold the person responsible for them would be mistaken, and since praise and blame go beyond the action to the intention they presuppose responsibility.) It would also be mistaken to evaluate the agent's character on the basis of an action that is unintentional, for an unintentional action would not be
a reflection of the agent's character or intentions (barring negligence, which reveals the extent to which the agent values certain results by indicating the level of care the agent maintains to achieve them), and any assessment of such action would have no bearing on an assessment or evaluation of the agent's character. If an agent is responsible for an action which is judged to have good effects or which surpasses the expectations placed on the agent in some way, the action and the agent are praised. If the results are judged to be bad, or fall below the standard of behaviour placed on the agent in some way, then the action is criticized and the agent is blamed. To praise is then both to laud an action and to approve, esteem or compliment the agent; to blame is then to criticize an action and to reprove, censure, rebuke or criticize the agent. When the standards against which the action is measured are moral (in that they refer to moral values, rather than, for example, aesthetic ones), assessments of the morality of the action become linked to evaluations of character. The assessment of the moral worth of the action then appears to entail the evaluation of the moral worth of the agent.

6. Character and moral worth

The force of Hospers' argument lies in his fixing a limit on praise and blame by pointing out that one is never justified in assessing the moral worth of the agent or in praising or blaming agents for what they are, rather than for
what they do. An action may have been within the agent's control, and so reflect the agent's character, but ultimately the agent's character was produced by factors beyond the agent's control (such as heredity and environment) and so is ultimately a matter of good or bad luck. (Of course, insofar as one shapes character by encouraging some actions through praise and discouraging others through blame one directs praise and blame toward the agent's character, but that is not the same thing as praising or blaming the agent for being what the agent cannot but be.) The agent who acts morally cannot take credit for the heredity and upbringing that made the agent that way; nor is the immoral agent to be blamed for the factors that produced that agent's character. Blameworthiness and praiseworthiness do not follow from responsibility insofar as they are applied to the agent's character rather than to the agent's actions, since it is inappropriate to hold an agent responsible for something which is not subject to the agent's control, such as character, and insofar as praise and blame go beyond the action to the agent (in the manner indicated) it is responsibility for character that praise and blame require, rather than responsibility for the action. A moral evaluation of acts for which the agent is responsible will allow one to make a moral evaluation of the agent's character only in that one may say of the agent that the agent does or does not act morally, and one may or may not esteem such behaviour. (Thus, in saying the agent is "morally good" one says that the agent acts morally, but to say that the agent acts morally because the agent is
It does not allow one to praise or blame the agent for having the character that is such that the agent acts morally or immorally. One can assess the moral worth of an action or of behaviour, and so praise or dispraise it, but one may not assess the moral worth of an agent. To praise or blame the agent, rather than merely esteeming or condemning the act or pattern of behaviour, is to assess the moral worth of the agent as well as to ascribe responsibility to the agent and to morally evaluate the act, and so is illegitimate. Insofar as praise and blame refer not simply to the act but to the attitude or intention of the agent which produced the act, they go beyond the act to the agent. Such a move is illegitimate because the agent cannot possibly be responsible for that agent's character in the way required by such an assessment (although it is legitimate to address praise and blame to the agent in order to shape character and future behaviour, as we shall discuss, which is not the same thing as praising or blaming the agent for being what the agent is, which is a matter over which the agent ultimately has no control). [24]

There are ways in which praise and blame may be justifiable, but before explaining how an ascription of responsibility together with a moral assessment of the act are all that is required (and that an evaluation of the agent's moral worth is useless), we might look at what sort of conditions would make praise or blame of the agent (for being what the agent is)
appropriate. If a person is praised or blamed for an action because of its intentions, then to deserve praise or blame that person must be responsible not only for the act but for that person's character, as it is that person's intentions and thus character which are being assessed. It is a logical impossibility for someone to be completely responsible for his or her character, for one cannot be the cause of oneself, as that would require that one pre-exist one's own existence. The idea that one can be completely responsible for one's own existence appears to be a hang-over from the Christian idea of autonomy of the soul. On this view, the soul is free from causal determination. One is responsible for it simply because it constitutes the essence of one's self. Though created by God, the soul is free. Whether it is good or bad is a matter of its own choice, and its choice is undetermined. Its choice has its origin only within itself, and it chooses freely and gratuitously, as did Adam and Eve. It does not choose because of its moral character, for in choosing good or evil it chooses its moral character, and as that choice emanates only from itself it alone is responsible for it, and totally responsible. It is the soul that determines how one acts, for no circumstances compel the soul to choose to act in one way rather than another. As one is responsible for one's soul, as it is the essence of one's self and so one is in a sense the cause of oneself as the soul is responsible for itself, one is responsible for one's character. One is thus deserving of praise or blame for the moral quality of one's
actions, as their goodness or badness depends on the goodness or badness of one's soul. Hence, one can morally evaluate the goodness or badness of a person on the basis of that person's acts, and so praise or blame that person. As persons are ultimately and totally responsible for their own actions, a person is thus deserving of reward or punishment simply because of that person's moral worth. This view then sees retribution as an appropriate response to moral badness and reward as a fitting response to goodness. A person can then be seen as simply morally superior or inferior, as the person is responsible for his or her own character. Good persons should be then (paradoxically) entitled to take pride in themselves; bad persons should feel a sense of guilt and sin. One can judge oneself in the same way as others can judge one. The view outlined here allows for a whole system of praise and blame and a notion of moral worth and deservingness based on a metaphysics of the soul which makes even less sense than Campbell's metaphysics of the self, embodying as it does the logical impossibility of the self being the cause of itself (which is another way of saying that the soul is responsible for itself and so one is responsible for one's soul). Moreover, the only support for this system is the impossible metaphysics. (Kant's noumenal self and other similar notions simply present the soul in Rationalist drag, and though Sartre states that consciousness cannot be the cause of itself, in speaking of freedom to choose oneself through an original project he seems to forget this.) If this Christian metaphysics of the soul is impossible
(as many Christians, particularly those who hold that salvation is a matter of divine grace, think it is), then so is any moral system that assesses the moral worth of the agent, since any such system must presume that the agent is responsible for his or her own character. Praise and blame, insofar as they involve this sort of character assessment, are then never legitimate.

7. Praise and blame as attitudes and as actions

It should be made clear that in speaking of praise and blame, one refers to an action and to an attitude. To praise or blame someone is to consider that person responsible for an action or attitude which falls above or below some standard that members of society are expected to meet. It may also be to consider that person's character as one that is good or bad in that it intends actions or exhibits attitudes which are good or bad (according to some standard). In this respect praise and blame are attitudes one holds toward the agent. To consider an agent worthy of praise or blame is at least to consider the agent as the source of some effect which is judged as good or ill, as when a person is said to be "at fault" in an accident or when one feels gratitude toward a crazy benefactor, to take two examples. In those cases, in fixing praiseworthiness or blameworthiness one is more interested in determining the agent's relation to some effect than in whether the effect was intentionally produced. If the agent is to be worthy of praise and blame at all, then the agent must be responsible for it in some way, such as neg-
ligence, even if the effect was unintentional. Naturally, the extent to which the agent is responsible for the effect is the extent to which the effect is intentional and the agent was acting rationally, and so the degree of strength in appropriate praise and blame is correlative with the degree of responsibility of the agent. The strongest sense of praise and blame refers to attitudes of the agent as evidenced in actions which, because of the agent's level of care in the performance of the action or because of the intention of the agent, is seen as meeting, surpassing or falling below a standard of human conduct and attitude. It is this sort of praise and blame that comes into play when we consider an agent "morally responsible," and the application of such moral standards to actions presumes that it was within the agent's power to have determined the effect of the action, and that application of those standards in that situation would not place unreasonable demands upon the agent. They are based on a practical concern for the controlling of certain effects through the control of their causes, and their appropriateness is contingent upon the degree to which those causes may be controlled. The verbal expression of those attitudes of praise and blame, or their embodiment in punishment and reward, constitute the actions of praising and blaming, and while the inappropriateness of the expression of the attitudes of praise and blame does not necessarily indicate the inappropriateness of those attitudes, the appropriateness of such attitudes is contingent on the possibility of some expression of them in
word or action being efficacious under at least some circumstances. For that reason, their appropriateness is contingent on the responsibility of the agent, as the causal efficacy of the action of praising or blaming requires that the agent have the agency to alter his or her actions due to rational considerations, such as the opinions of others and rewards and punishment.

What we have said about the illegitimacy of some kinds of praise and blame applies not only to the actions of praising and blaming, but to the attitudes which form the basis of those actions. Since a person cannot be ultimately and fully responsible for his or her character, one is not only wrong in expressing a valuation of that person's moral worth, but in even making that valuation. It is as unjust to presume to judge people as contemptible, evil, bad, despicable, admirable, holy, good or enlightened as it is to express that opinion. As if people had any say in their heredity and upbringing! Of course it is legitimate to consider a person's character and evaluate it when what one is referring to is a person's behaviour, which is to describe character in an evaluative way without ascribing responsibility for it, but it is wrong to go beyond that to praise or blame that person for being the sort of person who acts that way. That is, it is wrong to admire or despise the person, rather than the person's behaviour, for being what that person is, since what the person is results from causal factors beyond that person's control. What this rules out is not the judgment that so-and-so is a criminal or so-and-so is a saint, but the attitudes of moral
condemnation or admiration that go with those judgments, carrying with them as they do the God-like presumption of moral superiority (or, less likely, inferiority) on the part of the person making the assessment. It rules out the idea that one performs good actions because one is good, or that one performs bad actions because one is evil, for to perform good actions is what it is to be good, and so being good is not the cause of performing good actions, as it is the same thing under a different description. What causes one to perform good actions, or bad ones, are things like heredity and upbringing, matters over which one has no control and for which one is not responsible. The idea of there being evil persons is ruled out altogether, for to say someone is evil is not merely another way of saying that person performs bad actions, but it is to claim that those bad actions result from that person's nature or essence, whereas if determinism is true, the reason the person performs bad actions is that the person had a heredity and upbringing that caused him or her to act that way, not because the person is innately and by nature evil. It is these attitudes of moral judgment that determinism makes illegitimate. As will be discussed later, those attitudes are often involved in praise and blame, but that does not entail that the responsibility of the agent for an action justifies their being held. Our claim is then not about the meaning of "praise" and "blame," for even if attitudes of moral indignation and moral worship are components of attitudes referred to as "praise" and "blame," they should not be. This is the force of Hospers' point. Assessments of moral worth are
unjust because they presume a freedom of self-determination that does not exist.

In fact, indignation and worship have more the character of sentiments such as affection and hatred than they resemble moral judgments. To praise or blame a person for a character for which that person is not responsible is analogous to praising or blaming a person for an ability (which is a product of innate and environmental factors, such as training, over which the person has no control) or beauty (which, except for environmental influences on health, is entirely an innate trait). Essentially, in expressing admiration or contempt for the quality of the person's traits in respect to those categories one is making an aesthetic, rather than a moral, judgment. The person is judged by some standard of arete, be it virtue or beauty. But while possession of arete may call for praise, lack of it does not call for blame. Praise is an ambiguous concept in that it can be used as an aesthetic and as a moral judgment; blame is not. Here the praise and blame dichotomy breaks down, indicating that one is not making a moral assessment when one assesses a trait, as one is not assessing that for which the person is responsible. To praise beauty seems fine, but one does not "blame" ugliness. To do so would simply be inappropriate, as well as cruel. The term which corresponds to "praise" here is not "blame," but such words as "dislike," "abhor," "criticize" and so on, words which indicate a lack of appreciation for or a lack of attraction to some object. And it is worth noting that the presence of
such attitudes indicates that the person is being evaluated simply as an object for the person making the evaluation, not as a subject with its own system of values, likes, dislikes and needs, in that the person is assessed relative to the needs, values and preferences of the person making the judgment. For that reason the judgment is aesthetic; it is, like preferring chocolate to vanilla, a matter of taste. While one may admire a beautiful person, that person can take no credit for being beautiful, and an ugly person is not to blame or at fault for being ugly. The same is true of admiring or despising someone for that person's character or moral qualities. One may admire a person's character, but the person can take no more credit for having such a character than one can take for being beautiful, talented, athletic or a male caucasian over five feet tall. Similarly, to blame a person for that person's moral character is wrong, in that it is inappropriate, even if to disapprove of that person's character or behaviour, and even to passionately dislike it, may not be inappropriate. Indignation and admiration are passions, rather than judgments, and their role in the moral sphere will be discussed more later. It is enough to note here that the nature of admiration and indignation indicates that while like and dislike of character may be appropriate and indeed inevitable, praise (in its moral rather than its aesthetic sense) and blame are not.

This, of course, does not rule out the legitimacy of some attitudes of praise and blame. If in ascribing blame one is merely considering the agent as the cause of a certain effect
and as being in some way responsible for it (through intent or negligence) while considering that effect as undesirable, bad or lamentable, then that is justifiable. Or if one is simply saying that actions over which the agent had control did not meet certain (perhaps moral) standards, then that too is justifiable (though such an attitude would be more properly construed as criticism, rather than blame). But to move from those descriptive assessments to a condemnation of the agent is to condemn the agent for what the agent is, rather than simply criticize the agent's actions or behaviour, and is illegitimate, as it presupposes that the agent was free to have been otherwise than what, given determinism, the agent was and must have been, and it makes no difference whether such an assessment is thought, felt, spoken or acted upon. The same goes, mutatis mutandis, for praise.

8. Praise and blame and agency

So much for attitudes of praise and blame. What is the relation between a person's responsibility for an action and its moral worth as reflected in the practice of holding person's responsible? We have said that the appropriateness of attitudes of praise and blame is contingent upon the possibility of the efficacy of the expression of such attitudes. We can now provide a rationale for the connection between the appropriateness of attitudes of praise and blame and the ascription of responsibility. In so doing, the relation between the appropriateness of attitudes of praise and blame and the efficacy of the expression of those attitudes may be brought out.
As has been pointed out, to express the belief that one holds a person responsible is to indicate to that person that it was within that person's power to perform or not perform an act in the particular situation in which it was (or was not) performed. Moreover, it is to point out in what way the person had control of the action (or could have had control of it) and so is a way of making the person more aware of that person's agency. The person's agency and freedom are thereby increased, as the person is made more aware of the possibilities open to him or her, and is then more able to consciously direct actions toward his or her purposes. This enables the person to exercise greater control over actions in the future, as awareness of agency and the possibilities of choice allows for greater use of one's powers toward ends of which one is aware, and thus allows for greater freedom. To hold someone responsible is distinct from praising or blaming that person. It is not to make an assessment of the worth of the action or of the agent, but is an evaluation of the agent's freedom. It is a recognition of that freedom that when expressed allows the agent to act more freely and thus with greater control and responsibility in the future. One can, quite independently, assess the moral worth of an action. When one expresses an assessment of an action for which the agent is responsible as being good or bad, one points out to the agent that agent's responsibility for that good or bad action. But in so doing one is merely pointing out to the agent the agent's capacity for avoiding or repeating
that action and the fact that such actions are looked upon with favour or disfavour by some group or individual. If the agent is aware of his or her capacity for performing or not performing the action, and was aware of it at the time of the action, then the agent is forced to acknowledge that he or she desired its outcome. If this is the case, one is then in the position to assess the character of the agent as the sort of person who desires or intends certain things, which one might regard as being either good or bad. One is not in a position to pass judgment on the agent's moral worth, for that is to go beyond the ascription of agency and assessment of character to an ascription of responsibility for one's character, a move that, given determinism, is illegitimate.

If there is anything to the practice of praising and blaming, it is in the efficiency of the expression of praise and blame in controlling behaviour. To praise or blame is to assess the value of an action and to ascribe responsibility to an agent. In ascribing a value to the action, one indicates whether or not it is acceptable and what steps will be taken to control the frequency of its occurrence. That is, one expresses approbation or disapprobation together with an indication of what consequences an agent may expect from others as an incentive to repeat or abstain from such actions. The practice, as Schlick has pointed out, is intimately connected with the ideas of reward and punishment.25 One does not reward or punish because of the agent's moral worth; one rewards or punishes because of the worth of the act and because one wants to encourage or discourage the performance of such actions.
(This is compatible with the idea that punishment should be
treatment or rehabilitation of the offender, as even in the
form of treatment, punishment contains an element of censure
of the action and the agent, which the agent may wish to avoid,
and also involves some interference with the liberty of the
agent. Punishment need not be painful to be a deterrent.)

Praise and blame as actions are verbal forms of reward
and punishment which appeal to the agent's desire to be
esteemed or liked by his or her fellows. This esteem may
relate to the agent's motives or the agent's skills. Praise
and blame when applied to the agent's intentions, and hence
the agent's character, then appeal to the agent's desire to be
moral, or to at least appear moral to others in cases where
morality is not internalized. Praise and blame applied to
the agent's method of action or the manner of performance
appeal to the agent's desire to do well or to have his or her
skills recognized and appreciated. This is a rough sketch:
the psychological mechanisms of the effects of praise and
blame on motives and intentions is a complex matter. But it
is sufficient to note here that praise encourages actions and
blame discourages actions by appealing to an aspect of the
agent's character which may stand in opposition to those
aspects of the agent's character which would otherwise lead
the agent to perform or not perform a certain action. By
appealing to a sense of duty or desire for esteem one over-
comes one agent's indolence; by appealing to consideration of
others or a desire to live in society one restrains an agent
from committing an offence. In order for praise and blame to
be effective, they must appeal to some desire or motive of
the agent that stands in opposition to desires and motives that
the actions of praising and blaming try to counter-balance or
overcome. One thus attempts in praising and blaming to modify
the intentions of the agent by pointing out to the agent that
the consequence of certain actions is the approbation of
disapprobation of the agent by others. An agent whose actions
can be altered by praising and blaming is thus one who is
concerned with how he or she is thought of or regarded by
others. As praise and blame are incentives, they are not
coercive as they require that the agent be able to act on
incentives, and thus require that the agent control his or
her actions through rational and conscious choice. Incentives
are thus a way of appealing to the freedom of the agent, rather
than a limitation on that freedom. They encourage the agent
to take control of certain actions in view of certain con­
sequences and thus encourage the agent to recognize freedom and
agency, and so to be responsible. Praise and blame thus appeal
to the agent's freedom and to the agent's responsibility.

For that reason, praise and blame, like other threats and
inducements, are applicable only to responsible persons.
Responsible persons are the only ones who may alter their
intentions and actions due to a consideration of the con­
sequences, because they are in conscious control of their
actions. Praise and blame are like any other consequence
of an action except that they are applied by other persons
in order to encourage or discourage actions by making it
known how those persons regard such actions and what steps they will take to prevent or promote them. More generally, one promotes or deters behaviour by offering certain consequences as threats and inducements, and such consequences may only be applied to responsible agents as they are the only agents able to alter their actions by considering the consequences of those actions. The reward may be praise or esteem or something more tangible, depending on the value of the action; the punishment may be criticism or something more drastic, depending on how important it is to deter the action and what measures need be taken for doing so. The extent of the reward or punishment has nothing to do with the agent's moral worth; it is based on the value of the action (and hence on the desirability of that action being promoted or deterred). The form of the incentive or deterrent will depend on what is likely to control behaviour in the desired manner: if the agent is one concerned with the opinions of others, praise and blame may be sufficient; in other cases, more material measures (such as monetary reward or the deprivation of liberty) may be necessary. But in any case of incentives or deterrents, one appeals to the agent's ability to control intentions and actions in view of their consequences, and thus one appeals to the agent's freedom. This is very unlike physical constraint, mere imprisonment or forced actions, where no such appeal to intentions exists and no choice is allowed the agent. It is also different from behaviour modification techniques such as electro-shock therapy, which instead of increasing agency and appealing to the agent's freedom, remove the freedom of the agent
by making certain choices compulsive, and hence irrational (because they are not within the conscious control of the agent) even if they appear to be in the agent's best interests. At best, such practices merely substitute one compulsion for another. By taking away freedom, such practices also remove responsibility. (A good literary illustration of this can be found in Burgess' *A Clockwork Orange.*) The practice of praising and blaming, like other forms of reward and punishment, is a unique form of behaviour control in that it actually seeks to increase, rather than limit, the freedom and responsibility of the agent.

One judges whether to reward or punish, praise or blame on the same basis as one judges whether to perform any other action: on the basis of its effectiveness, of its likelihood of producing a desired outcome, namely, the control of the behaviour of the person being praised or blamed. Its efficacy is dependent on several necessary conditions. It has already been argued that the agent must have been free and rational at the time of the action. But the agent must also be free and rational at the time of the reward or punishment, praise or blame if such measures are to be effective, for even if the agent was rational at the time of the action, the purpose of incentives and deterrents is to control future actions by appealing to the agent's ability to control or later his or her intentions, and that is impossible if the agent is no longer rational and free. If after committing a murder with full agency an agent goes insane, it would be inappropriate to
treat the agent as one would treat a sane agent, for the simple reason that the agent is no longer consciously controlling his or her actions, and so will not be affected by a consideration of the consequences of those actions. An agent once in possession of the freedom to realize intentions but is now lacking it is similarly not affected by considerations of the consequences of actions, as such an agent is either simply not able to act or is restricted in doing so in such a way that the agent is not capable of the actions being encouraged or deterred. The robber who is paralyzed by an injury sustained in the course of committing a crime is no longer capable of breaking the law, or, for that matter, of following it: freedom and responsibility vanish together. It may be objected that incentives and deterrents are supposed to promote or deter actions on the part of other rational or free agents as well, and failure to follow through on threats or promises of consequences will result in those consequences not being taken seriously. This apparent difficulty is, however, illusory. In fact, loss of rationality and freedom are in themselves deterrents, and so will themselves affect the behaviour of others. Like other deterrents, they may not be totally effective, but certainly it is a different thing to commit a crime knowing one will not be punished if one goes insane in the meantime than to "get away with murder." One who loses rationality or freedom is not "getting away" at all. Certainly, if the alternative consequences of an action are insanity, loss of freedom to translate thought into action, or punishment by others, such an action will not appear very
attractive. The conditions which would make application of punishment or blame ineffective are themselves powerful deterrents, and so the application of further deterrents would be ineffective and superfluous (and hence an unjustifiable infliction of suffering).

Like other actions, praising or blaming may be just or unjust. It would be unjust to praise or blame a person not responsible for an action merely to give an example to others because the selection of that person would be arbitrary. Moreover, no one would want to live under a system of arbitrary reward and punishment, for under such a system such consequences could neither be achieved nor avoided through intentional action, as they could befall anyone, and so the purpose of such incentives and deterrents (the control of behaviour through an appeal to intentions) would then be lost. Under such a system, rewards and punishments would not function as incentives or deterrents, and so would not really be rewards or punishments, but some other practice whereby material gain or suffering or approbation or disapprobation are allocated in some sort of arbitrary way. Such material measures without the responsibility of the agent would not be reward or punishment, and such opinions would not be praise or blame. Not only would such measures be unjust, they would be entirely inappropriate, for in praising and blaming one refers to what the agent intentionally did, and if the act did not reflect the agent's intentions, or if no act was performed and the agent was innocent (so to speak) of any action, one would be ascribing to the agent intentions (or even actions) that did
not exist. Though praise and blame do not follow from responsibility, they require it, since to praise or blame is to make a claim about the agent's responsibility for an action and hence about the agent's intentions.

Praise and blame also require the responsibility of the person being praised or blamed for the action because they are addressed to the agent as well as to others, and the agent can only be encouraged to repeat or refrain from certain actions if the agent has the agency to perform those actions intentionally. Since praise and blame appeal to the agent's agency, as they are directed at controlling or altering the agent's actions by controlling or altering the agent's intentions or method of action, one does not praise or blame unintentional acts. To do so would be to encourage such acts, and to encourage or discourage such acts is pointless, as the agent is not in control of those actions, and so such encouragement or discouragement can have no good effect. One may, however, criticize an unintentional act for its negligence, in which case one blames the agent for the agent's method of action or the manner of performance of the action, matters over which (assuming the agent was negligent) the agent has control and which are therefore subject to control or alteration from praise and blame. Similarly, an act can be praised not only for its intentions, but for its manner of execution, which encourages that method of action. But to praise or blame the agent for actions or aspects of actions not within the agent's control is pointless because it is inefficacious and unjust because it wrongly ascribes an agency to the agent which the agent lacked. In
ascribing to the agent agency that the agent lacked, and hence intentions or attitudes the agent did not possess, one misdescribes the agent's behaviour. Praise and blame can only be effective and just when the agent is in control of the action or aspect of an action being praised or blamed.

To say then that an agent "deserves" praise or blame is to indicate that praise or blame would be appropriate, and praise and blame are appropriate when the agent was in control of the action or behaviour being assessed. Desert is then not a matter of moral worth and hence of what certain moral types deserve; it is a matter of what person is the appropriate object of a method of controlling behaviour, such as praise or blame, and of whether that person was responsible for the action and therefore an appropriate object of such measures. While the notion of desert is backward-looking in that it is related to an assessment of the value and cause of a past action and of whether the agent was responsible for that action, it is forward-looking in that it is bound up in the appropriateness (i.e., the likely efficacy and the necessity) of praising or blaming the agent. It is backward-looking in that the question of whether to praise or blame (or not) is dependent on whether the agent was responsible for the action, that is, on whether the agent was in control of the action. It is forward-looking in that an answer to that question determines the probable efficacy of praise and blame, or of any other form of threat or inducement, in controlling such actions.

Responsibility is only a necessary condition of the appropriateness of praising and blaming, as it is necessary for the
efficacy of those practices but not sufficient. To deserve praise or blame is not only to be responsible for an action, but to be in a position where receipt of praise and blame may affect one's future actions. One may go so far as to say that one is deserving of praise or blame when the good effects of being praised or blamed are likely to outweigh any possible bad effects. (One may note that the efficacy, and thus the appropriateness, of praise and blame is an empirical matter. It is possible that psychological investigation will show that in fact blame, like guilt, is more debilitating than effective, or that praise leads to egotism. This thesis allows for the possibility that praise or blame may be inefficacious means of controlling behaviour; our object here is merely to show what conditions must necessarily obtain for either praise or blame to be effective.) A person may be the appropriate object of an attitude of praise or blame, and in that sense be "praiseworthy" or blameworthy," but expression of that attitude may be inappropriate or unjust under some circumstances. It may be pointless or even harmful to praise or blame even when a person is responsible for an act. Virtue is its own reward often enough, and recognition of agency and freedom and of the value of the act may be present to the agent and so need not be brought to the agent's attention (though to do so would do no harm, and to express one's admiration or gratitude may benefit the agent and so serve as an additional, if unnecessary, way of promoting the action praised). In other cases, where a person achieves a good and profits personally from such an achievement, praise is unnecessary, and, in cases where personal
profit was the motivation and where the praise refers to the intentions or attitudes of the agent, it is misguided in that it incorrectly ascribes worthy intentions or motives to the agent which the agent did not possess. Similarly, evil can be its own punishment, and the recognition that one has done wrong and had the power to avoid doing so is often punishment enough, in which case further punishment (such as blame) would be a pointless and cruel infliction of suffering. The point of praise and blame is to guide actions by pointing out to agents the values of certain actions and their power to perform or not perform actions, that is, their own freedom and agency. There may be cases where praise and blame are not necessary because the agent is aware of agency and freedom and of the value of the action, and the consequence of the action is in itself a sufficient deterrent or incentive, making any consequences deliberately imposed by others because of the value of the act unnecessary. To praise one already content with the consequence of the action does little (though there may be cases where one is content only if the consequence of the action is praise), and to blame someone despondent over an action is petty and cruel, as well as useless (and it is also to assume a presumptious attitude of moral judgment of the agent's worth, and we have seen that such assessments are never legitimate).

9. **Reactive attitudes**

There is, however, an element of praise and blame which needs to be more fully explored. It is what has been alluded to as the aesthetic judgment of character and actions contained
in some forms of approbation and disapprobation. The basis of such judgments are what P.F. Strawson has termed "reactive attitudes." Insofar as praise and blame are expressions of attitudes of like or dislike toward the agent, they are more a function of the attitudes and expectations involved in interpersonal and social relationships than a consequence of the practice of holding persons responsible. They are related to the concept of responsibility only insofar as the responsibility of the agent is a necessary condition of the appropriateness of reactive attitudes being directed toward the agent.

We respond with like or dislike, aversion or attraction, to different events and states of affairs. Our feelings of like and dislike are directed toward actions, among other things. But these feelings are directed in some cases not at the actions themselves, but express a like or dislike of the intention or attitude of the agent the action reveals. The intentions and attitudes toward us and toward humans in general is a matter of prime importance to us as individuals and as social beings. It is important to us to be able to both recognize and to evaluate those intentions and attitudes as good, favourable or sympathetic on the one hand, or as malevolent, contemptuous or indifferent on the other. We must know in some measure what the intentions or attitudes of a person are in order to know how to interact with that person and whether or not one wishes to enter into a relationship, be it personal, social, political or commercial, with that individual. Awareness of the attitudes and intentions of others and concern for those attitudes and intentions thus profoundly affects our own attitudes, intentions and ways of life.
The reason for this is that certain relations between persons or between an individual and society require the presence of certain attitudes among the participants. The attitude required and the intensity or degree to which it is held depends on the relationship. Most generally, social life depends on certain attitudes of goodwill, compassion, caring or concern among the participants. Morality is essentially a regard for the interests, rights and feelings of other sentient beings. One who is incapable of such other-regarding attitudes is not a moral agent and is therefore incapable of moral freedom. (It is, of course, one thing to be immoral and another to be incapable of being moral.) Relations with such a person would then not involve treating the person as one capable of moral freedom and responsibility, but would rather involve an effort to control or alter that person's behaviour by appeal to other measures than moral sense, or a regard for others, as that sense is lacking in that individual. (The immoral agent may possess the sense, but it is not activated or dominant enough, in which case the purpose of holding that agent morally responsible would be to develop or activate that sense or to tie it to some other concern, such as a desire to enjoy the freedoms accorded only to morally responsible individuals.) In dealing with a person incapable of holding other-regarding attitudes, one, in a way, would be forced to treat that person as being somewhat less than a person, for it would not be possible to appeal to that person's freedom to control the moral qualities of that person's acts, as that person lacks such freedom. (Sociopaths are individuals of this type.) In
consequence, any relation with that person requiring moral freedom and responsibility would be impossible. Such a person would be subject to neither praise or blame, as such a person is incapable of holding the attitudes to which praise and blame (as assessments of attitudes) refer. On the other hand, one possessing the ability to have the appropriate other-regarding attitudes would be expected, as a condition of that person's participation in society, to possess those attitudes. When that ability is present in a person, we take pleasure in the fulfillment of our expectations regarding the attitudes the person should possess, and feel distress or hostility when those expectations are not realized. A person who fulfills our expectations by exhibiting appropriate other-regarding attitudes is a welcome participant in society; a person is perceived as noxious, and noxious due to choice (and hence evil), and so is abhorred or resented if such a lack of other-regarding attitudes is characteristic, and mistrusted as a participant in society en tous cas. Social relations are based on trust, and trust that others will regard persons other than themselves favourably; betrayal of this trust is a threat to social relations and to their participants, and the natural response to a threat is to be wary of it and to take steps to defend oneself against it. In cases where the threat is a person, one appeals to that person's interests, if that person is free, and to that person's interests in participating as a free and responsible individual in society, if that person is morally free in the sense that the person is capable of holding the appropriate
other-regarding attitudes. If the person does not exhibit
the appropriate other-regarding attitudes, that person is
excluded from certain sorts of social relationships until such
time as those other-regarding attitudes are evidenced, as
a person lacking those attitudes cannot fulfill the expectations
required of that person in those relationships. The same
basic reasoning applies to other relationships; in commercial
transactions, some honesty or responsibility is required; in
interpersonal relationships, the expectations on the attitudes
of the participants are much greater, and will include things
such as certain kinds of trust, love and loyalty. The necessity
then, of certain attitudes being present in the participants in
relationships entails that where a relationship exists, certain
expectations will be held of any individual participant, and
violation of those expectations will be seen as a threat to
the relationship and to its other participant(s).

Because of the expectations we place on others with regard
to the attitudes they hold toward us, an evaluation of the
agent's attitudes and intentions may be accompanied by an
emotional response based primarily on feelings of like and dis-
like. As the possibility of certain positive emotional
attitudes toward another is dependent on the presence of certain
attitudes in another, lack of those attitudes in another may
rule out certain kinds of emotional responses, and hence
certain kinds of relationships. In cases where attitudes of
a certain sort are expected to be held by another (as, for
example, when the other is a participant in some social or
personal relationship with one) and are absent, one's emotional
response may be based on feelings of aversion and dislike (based primarily on the perception of a hurt or potential hurt due to the other). The kinds of emotional responses one can have toward persons thus depends on the possibility of persons holding certain attitudes necessary for the appropriateness of such emotional responses and on whether, when such a possibility is in fact present, a person does in fact possess those attitudes. Any relationship hence requires that the participants hold expectations of each other, and to ask whether one should subject others to expectations is then to ask whether one should have relations with others.

The emotional response to an assessment of an agent's attitudes and intentions is often a component of attitudes of praise and blame and the expression of those attitudes. Feelings such as gratitude and resentment, which Strawson places under the general rubric of "reactive attitudes," are a product of one's assessment of the agent's intentions, rather than of the act itself. Specifically, reactive attitudes are a reflection of what one believes were the attitudes of the agent toward oneself or toward another in acting in a certain way. It is not so much the result of the action as the attitude which the action reveals or expresses that matters. Reactive attitudes are thus a reaction to an assessment of the agent's character, and are primarily directed toward the agent, rather than the action. They are directed toward the action only insofar as it reveals the character of the agent, and hence are ultimately directed toward the agent.
Certain forms of action are by nature inappropriate objects of reactive attitudes. Actions which are unintentional do not call forth an emotional response toward the agent simply because the action did not reflect the agent's intentions or attitudes, and it is those intentions and attitudes to which the reactive attitudes are a response. Negligent actions may appear to be an expectation since they do call forth a negative response, but are not so as they reveal the attitude of the agent toward oneself or others by indicating the pains the agent is prepared to take to achieve or avoid certain states of affairs that the agent believes (or is in a position to know) will affect others, or by indicating the agent's level of concern for others simply in view of the fact that the agent has or has not considered the effect such actions (or lack of care in their performance) will have on others. In the case of unintentional action, it is the fact that the action does not reflect the agent's intentions which makes the agent an inappropriate object of reactive attitudes based on that action, except in cases of negligence, where reactive attitudes directed toward the agent are based on the supposition that the action reflected an intention of the agent would be inappropriate, but where reactive attitudes directed toward the attitudes of the agent evidenced by the agent's negligence would not. Where reactive attitudes are suspended or are inappropriate due to the unintentional nature of the act, the suspension or the withholding of the reactive attitudes in no way implies that the agent is not rational and free in the manner required for an agent
to be responsible. So while the agent is excused because
the action was unintentional, and reactive attitudes are
therefore withheld, the agent is still regarded as being
subject to the expectations and demands that are part of
interpersonal and social relationships concerning the attitudes
of a member of such a relationship toward others. Reactive
attitudes are suspended since the fact of the unintentional
act is not incompatible with the possession by the agent of
an attitude or intention contrary to or incompatible with the
outcome of the act. Where an unintentional act is beneficial,
the act says nothing about the goodwill or virtue of the agent,
and so gratitude would be out of place. Or when the action is
harmful, the reason that the reactive attitude of resentment
would be inappropriate is that, as Strawson says, "the fact
of the injury was not in this case incompatible with ... the
agent's attitudes and intentions being just what we demand
they should be." In other words, not only would it be
reasonable to hold expectations that are part of full inter-
personal and social relationships with the agent, but the
agent has not failed to fulfill those expectations, since
the agent's actions were not a reflection of the agent's
attitudes and thus did not indicate that the agent lacks the
attitudes which we would expect the agent to hold. The inapprop-
riateness of directing reactive attitudes toward an agent on
the basis of the agent's unintentional act then says nothing
about the character of the agent or about whether the agent
is a possible object of reactive attitudes or a possible
participant in social and interpersonal relations.
There may be cases, however, where reactive attitudes may be inappropriate because the expectations of attitudes and behaviour of the agent involved in full interpersonal and social relationships are unreasonable. The expectations of attitudes and behaviour placed on others in a context of social or interpersonal relationships are unreasonable in two sorts of cases: those in which responsible agents are placed in extraordinary circumstances, and so their choices reflect the drastic or stressful nature of the situation, and those in which the agent lacks the rationality necessary to be able to possess the attitudes toward others necessary for full interpersonal and social relationships.

In our discussion of some actions which have been mistakenly called "involuntary," we noted that under circumstances where the choice with which the agent is presented involves some choices valued so highly by the agent that we could not reasonably expect the agent to sacrifice it or not act on it, the agent's actions are justifiable or permissible because of the value choice on which they are based. The agent is not excusable, because the agent possessed full agency and the action resulted from a rational choice of the agent, and so the agent is fully responsible for the action. Yet when drastic choices involving values held very strongly by the agent are presented to the agent due to extraordinary circumstances, the expectations we would have of the agent under ordinary circumstances (where such high values are not at stake) would not apply, as it would be unreasonable to expect the agent not to act upon certain important or highly ranked
values (and any rule asking that of agent's would be ineffective for that reason). In effect, because one cannot reasonably expect agents acting under such extraordinary circumstances to behave as they would under normal circumstances, those agents' obligations would not be the same as they would be under ordinary circumstances, if any obligations would even exist under extraordinary circumstances. If an agent does succeed in sacrificing a highly valued goal (such as self-preservation), then such acts are indeed estimable, as they far surpass what we may reasonably expect agents to do. It is not that the expectations we would have of the agent in normal circumstances are met and then surpassed; by achieving a good we would expect of the agent under ordinary circumstances, the agent achieves a good we would not expect the agent to, and so, far from merely surpassing our expectations, the agent achieves a value when no expectations of the agent to do so are present. The value of the act derives from the fact that the agent went beyond the whole network of normal expectations of other-regarding attitudes and behaviour. If the agent does, however, choose some very high value (such as self-preservation), though such an action is not excusable (since the agent is in full control of it and so fully responsible for it), it is pardonable, as the expectations one would normally have of an agent do not apply in that situation, and so the act does not violate the expectations of goodwill that form a basis for full interpersonal and social relationships. Where one cannot reasonably expect the agent to do otherwise, one cannot reasonably blame or resent the agent for the action.
On the other hand, if an agent achieves a good because non-performance of that act would result in the sacrifice of a very high value, it would be inappropriate to praise or feel gratitude toward the agent for doing what one could not have reasonably expected the agent for doing what one could not have reasonably expected the agent not to have done. In such cases, it is not correct to say that the agent has fulfilled our expectations, as the motivation of the action was the preservation of some highly valued goal, rather than some other-regarding attitude, and our expectations refer not only to behaviour but to the attitudes which produce it.

Because of the basic nature of highly valued goals such as self-preservation, and because one cannot reasonably expect an agent not to act on those values, actions based on highly valued goals, while they are revealing in that they illuminate the value system of the agent, in a way say little about the agent in that they only indicate that the agent is not unlike most people in that the agent holds certain values so highly that it would be extremely unlikely for the agent not to act on them. Hence, reactive attitudes directed toward agents on the basis of actions that take place in circumstances where the agent acts on a value one could not reasonably expect the agent to sacrifice are inappropriate because the expectations that one could reasonably hold of the agent under ordinary circumstances cannot be reasonably held in circumstances where the choices presented are far more drastic. The only expectation would be positive reactive attitudes toward heroic acts or acts of great self-sacrifice, but those
feelings would be present precisely because the agent has achieved some good in the absence of any expectations of the agent to do so. In surpassing our expectations by achieving in extraordinary circumstances that we can reasonably expect of agents only in ordinary circumstances, the agent earns our gratitude. The existence of negative reactive attitudes for a failure to go beyond what we can reasonably expect of agents would be unreasonable and hence inappropriate, and the existence of positive reactive attitudes toward an agent who did good because not doing so would result in drastic sacrifices on the part of the agent would be inappropriate and in a sense mistaken. Reactive attitudes and the expectations on which they are based would be suspended, however, due to the nature of the circumstances in these cases, and this suspension would be fully compatible with the agent being responsible and so subject to the expectations held in full interpersonal and social relationships and so to the reactive attitudes on which they are based. The agent in such situations is still responsible and free; it is our expectations which differ or are absent, as the case may be.

Circumstances may make reactive attitudes inappropriate for a different reason, namely, that the stressful nature of some situations causes the agent to behave irrationally. Under such circumstances the agent is not responsible and therefore not subject to the expectations we place on responsible agents or to the reactive attitudes resulting from those expectations. Stress may, moreover, cause the agent to act "out of character," so that the agent's actions are not truly reflective of the
agent's real attitudes, which would be operative and detectable in circumstances where the agent was rational. When under stress, considerations normally important to the agent may have no effect, making the agent's action a product of a momentary aberration in attitudes due to unusually stressful circumstances rather than to the attitudes the agent would exhibit when rational. The expectations one would hold of the agent when the agent is rational and the reactive attitudes that accompany them would thus be inappropriate because the agent is not responsible (as the agent is not rational) in those situations. Hence, the agent is not an appropriate object of the expectations and attitudes that are part of full interpersonal and social relationships because the agent, lacking control over actions, is temporarily incapable of engaging in such relationships (as the agent lacks the ability to fulfill the expectations which obtain in such relationships), due to irrationality brought on by stressful circumstances.

The second sort of case where the expectations of the agent which form the basis of reactive attitudes are unreasonable or inappropriate is when the agent is incapable of meeting those expectations due to a handicap such as chronic irrationality. Expectations are inappropriate in these cases because of a fact about the agent, rather than a fact about the circumstances. It is obvious that a physically handicapped agent cannot be expected to perform those actions that the handicap prevents the agent from performing. It should also be obvious that it is unreasonable to expect the irrational agent to have those attitudes of goodwill and prudence that the agent's
irrationality prevents the agent from having. Whereas the physically handicapped are not subject to certain expectations as to behaviour, they are, however, subject to expectations concerning what their attitudes toward others should be (barring any irrationalities, including those that may result from the handicap), and so are appropriate objects of the expectations and attitudes that are part of full interpersonal and social relationships. The mentally handicapped or irrational person is in a different situation. Such a person is incapable of holding the attitudes necessary to such relationships and which others would demand of a person in such a relationship. Excluded from certain interpersonal and social relationships (though not from therapeutic ones), mentally handicapped and irrational persons are therefore not subject to the expectations inherent in those relationships, and therefore their attitudes and intentions are not subject to reactive attitudes based on those expectations.

The appropriateness of reactive attitudes thus coincides with the agent's being responsible for the action in most cases. As reactive attitudes are directed toward the attitudes and intentions of the agent, it is necessary that the action reflect the agent's attitudes and intentions, and thus (barring negligence) must be intentional. An unintentional action reveals the attitudes of the agent only when it reveals negligence, and then the reactive attitude is directed toward the fact of negligence, rather than the unintended outcome of the act. Reactive attitudes are based on expectations of what attitudes agents should have in certain interpersonal and social relation-
ships. Where the expectations are unreasonable, then so are the reactive attitudes. An agent can be responsible and yet be an inappropriate object of ordinary reactive attitudes if the demands placed on the agent are inordinately great, and in these cases the appropriateness of reactive attitudes does not correspond with the responsibility of the agent. However, an agent not in conscious control of his or her intentions cannot be expected to make those intentions conform to what we expect or demand of responsible agents, and such an agent is neither responsible nor a fit object of reactive attitudes, and so here again responsibility and the appropriateness of reactive attitudes coincide.

So responsibility at most outlines where reactive attitudes could be appropriate. As we have already seen, sometimes the unreasonableness of the expectations on which ordinary reactive attitudes are based can make such attitudes inappropriate even when directed at responsible agents. In fact, reactive attitudes are more a function of our expectations of other people than of their responsibility. Responsibility of the agent is a necessary but not sufficient condition for the appropriateness of reactive attitudes. The existence of reactive attitudes is contingent upon our concern for the attitudes of another toward oneself and others, and upon the existence or a desire for the existence of some interpersonal or social relationship in which another participates and in which the other's attitude toward oneself and others would matter.

If the relationship in question is an interpersonal one between the agent and oneself, then if one is indifferent to
how the agent feels toward oneself, no feeling of gratitude or resentment could arise as a result of the agent's actions that would be directed toward the agent. (One may like or dislike what the agent does, but those feelings would not be extended from the action to the agent.) Lack of concern for how the agent feels about oneself in fact rules out many of the deeper kinds of interpersonal relationships existing with that agent. Perhaps complete indifference toward how any rational being regards one is unlikely, but we may say that (roughly) the intensity of the reactive feelings varies proportionately with the personal, emotional investment one has in the agent or in a specific kind of relationship with the agent. It will also vary proportionately to one's expectations of the agent. These two areas are related in that the deeper the relationship, and hence the greater the expectations involved, the more serious becomes the fulfillment of expectations that would be the sum total of a less personal or deep relationship. Thus, lack of fulfillment of an ordinary expectation, which one would hold of any member of society, on the part of an agent with whom one is considerably more deeply involved is a much greater violation of expectations than it would be in a less deep relationship. Alternatively put, the expectations of a less deep relationship are basic to any further expectations that would be part of a deeper relationship, and so the emotional reaction to violation of expectations is bound to be that much stronger. To think of it geometrically, in interpersonal relationships one is concerned with how the other person feels about one, and the
greater the concern, the greater the radius of the circle within which actions subject to reactive attitudes will fall, and the nearer to the centre of the circle, the stronger will be the reactive attitude. (Thus a lie told by a trusted friend is resented more than one told by a stranger, and the murder of a parent is found more loathsome than the killing of one to whom one is not related.)

The expectations involved in interpersonal relationships extend into surprisingly many spheres of life. Even in a relationship between a patient and a therapist, in which certain interpersonal attitudes are deliberately excluded for the sake of objectivity, some actions will still be the appropriate object of a reactive response. Certain co-operation is expected between patient and therapist; the therapist is expected to act in the patient's best interests, and the patient is expected to follow the therapist's guidance. If, for example, the therapist is never in at the time of the appointment, certain expectations inherent to the relationship are violated, and if either the therapist or the patient tries to kill the other, expectations of goodwill basic to any relationship whatsoever are violated. Violations of the expectations inherent to the relationship reveal attitudes (eg., of indifference or hostility) which, given the nature of the relationship, matter to the participant enough that their expression or revelation violates the participant's expectations and injures his or her feelings or person, since the feelings (eg., of affection or trust) invested in the relationship have been betrayed or denied.
Generally then, if one cares about or feels for a person or cares about a relationship with a person, then one will care about the attitude of the person toward oneself. Reactive attitudes are then based on what one would like the person's attitudes toward oneself to be, or what one expects them to be given the nature of the relationship, rather than on what action or attitude one is entitled to receive from the agent. The less personal investment one has in a relationship with the agent, the less important what one expects of the agent or would like the agent to do (or to be) will be.

Since the above is the case, reactive attitudes based on an action that reveals the agent's attitudes toward others, rather than oneself, differ in degree only, and not in kind, from reactive attitudes which arise in a context of interpersonal relations with the agent. When one has a reactive attitude to an agent because of the moral value of an action, one is reacting to an action which reveals whether or not the agent has certain attitudes which are generally expected to be held by members of society toward other members. Such a "moral" reaction can be experienced simultaneously with more personal feelings directed toward the agent, and the presence of more intense feelings is contingent on how close the agent or one affected by the agent's actions is to oneself. In other words, it is a matter of how much the agent's attitudes matter to oneself that allows for "moral" as well as personal reactive attitudes. The one form of reaction does not rule out the other. If an agent injures someone close to oneself, one's feeling may involve moral indignation, but it may also
involve feelings such as resentment and a desire for vengeance. Or, for example, if someone is cruel toward one, one's indignation may be moral, but if the agent is close to one and has violated more expectations than the general ones which count as moral, one's feeling may go beyond moral indignation, not because the act is not judged on its moral principles, but because there is an additional non-moral element due to the fact that expectations less basic than those which count as moral but which are inherent to the relationship have also been violated. Of course, there are expectations less basic than moral ones, violation or fulfillment of which would give rise only to non-moral reactive attitudes based on expectations which occur only in deeper interpersonal relationships. And one may experience such reactions in one's own relationship with the agent, or when the agent does or does not fulfill expectations in an interpersonal relationship with another, with whom one identifies or feels for in some way, even if it is just as a fellow human. The intensity of the reaction will depend on one's involvement with the agent or those affected by the agent's actions. If an agent is not close to one, one may feel moral indignation where one cared about the agent would feel resentment, or approval where another would have felt gratitude. Where another close to one is affected by an action of the agent, one may feel resentment rather than mere indignation. Where the agent is oneself, the degree to which one feels remorse or pride over an action will depend on how close those affected by one's actions are to oneself, and how basic are the expectations one has fulfilled or violated.
In order for a reactive attitude to be "moral" then, it must be based on expectations basic to the existence of interpersonal and social relationships. A reactive attitude based on basic or moral expectations can be accompanied by a non-moral reactive attitude when the relationship with the agent and the expectations involved in it go beyond basic social relationships. One can have non-moral reactive attitudes when what one reacts to is the fulfillment or violation of expectations not basic to social relationships but inherent in the relationship one has with the agent (or when one is close to one affected by the agent's actions, or when the person affected by the agent is in a relationship with the agent that is deeper than a mere social one and with whom one identifies or sympathizes in one way or another). One's reactions thus varies with the kind of expectations one has of the agent and with how much one cares about the agent's attitudes (which varies with how much one cares about the agent and with how much one cares about those, including oneself, who are affected by the agent).

To be concerned with how another person regards others and oneself is in a sense to regard that person as an object, as one thereby subjects the person to one's own values and desires by concerning oneself with how that person stands in relation to those desires and values; yet lack of concern with the agent's attitudes toward others and toward oneself indicates that one is interested in the agent only as an object and as more of an object, since one interested only in the agent's behaviour (and so in the effects the agent produces)
is not interested in the agent's subjectivity (or else one would be interested in the attitudes of which the behaviour is a manifestation). Lack of concern for the agent's attitudes toward others is in fact to not consider the agent as a moral or social being, and so to not consider the agent as a possible participant in social or interpersonal relationships (which require not only the moral freedom of the agent but the concern on the part of others with the moral quality of the agent's actions, and so with the agent's attitudes). In the vent of such lack of concern, one would have no reactive attitude toward the agent. If one does not care about others oneself, one will also not care about how an agent treats others except insofar as that reveals how an agent is likely to treat oneself, and as one has not yet been affected by the agent, no reactive attitudes would follow from the agent's actions, though one may consider how to interact with the agent on the basis of the agent's other-regarding attitudes (or the lack of them). Such attitudes on one's own part, however, would indicate a failure to regard others as anything save means to one's own ends, in which case one would oneself lack the attitudes necessary for certain social and interpersonal relationships.

The existence of reactive attitudes thus depends on some relationship with the agent. The occurrence of a negative reactive attitude may rule out future relationships, but that is because the expectations of some previous relationship had been violated, and not because reactive attitudes are forward-looking, as they are not. Different sorts of agents will be
capable of different degrees of relatedness, and so of meeting different sorts of expectations, so that not being in a relationship with an agent rules out reactive attitudes, while the incapacity of an agent to meet the expectations involved in certain relationships (due to irrationality, for example) rules out not only reactive attitudes toward the agent, but even the possibility of reactive attitudes (insofar as the possibility of relationships with the agent is ruled out). Future relationships may thus be ruled out in the absence of any reactive attitudes. Reactive attitudes depend on the actual existence of a relationship, not on the mere possibility of a future relationship, though an assessment of that possibility might follow from reactive attitudes. But an assessment of the possibility of a future relationship with the agent does not depend on the existence of a reactive attitude, and a reactive attitude may exist even when future relationships with the agent are impossible (due to the death of the agent, for example).

Reactive attitudes hence do not follow from the practice of judging and holding persons responsible, but from the existence of certain relationships in which the attitude of the agent matters to one for one reason or another. Caring about the attitude of the agent is inherent in any interpersonal or social relationship with the agent (to varying degrees). Lack of a reactive attitude toward an agent may reflect a failure to take the agent seriously or indicate that one is incapable of certain types of relationships since one is incapable of the appropriate emotional responses involved in
them: for example, a failure to resent a threat to one's life by a trusted friend may indicate a lack of caring about oneself, and a failure to feel gratitude for being rescued at the agent's peril indicates a lack of caring about the agent. Certain relationships thus require expectations that give rise to reactive attitudes, as they require one to care about the agent's attitudes and interests, and also require that one care about the attitudes and interests of those (such as oneself) affected by the agent's actions. For this reason, reactive attitudes are inherent in certain sorts of relationships.

From this it does not follow that the expression of reactive attitudes is also inherent in certain sorts of relationships. Whether or not to express a feeling, however appropriate that feeling may be, whether the feeling is positive or negative, is a pragmatic question, the answer to which depends on the likely effects of the expression of the feeling in certain circumstances. Anger may be appropriate when one is abused, but retaliation or a rude response may succeed in only making matters worse, in which case it would be far better to deal with one's anger in some other way than a direct expression of it aimed at the person who provoked it. There may be cases where one can have one's interests taken seriously only if one responds directly to the agent. Being a pragmatic matter, it is also an empirical matter. The same goes for any other reactive attitude; whether or not one should express the feeling is dependent on the situation, and discretion is the better part not only of valour.
This being so, it is obvious that it is perfectly possible, contra Strawson, to have a reactive attitude while at the same time objectively considering what one ought to do in regard to the agent in order to secure the repetition or cessation of the action which gave rise to the reactive attitude, or in order to maintain or charge an agent's attitude, since it is possible to at the same time have a feeling and to consider whether or how one should express it. Strawson is mistaken in believing that objective considerations of how to affect or control the behaviour of an agent are incompatible with reactive attitudes. Reactive attitudes may prevent one from considering the situation objectively if they are sufficiently strong that one's emotions prevent one from being rational, but that is a feature of emotion in general, and not of reactive attitudes in particular, and it is the strength of the feeling, not the presence of the reactive attitude per se, which rules out objectivity. And while it is true that one may consider how to affect or control the behaviour of an agent in cases where reactive attitudes are inappropriate, as for example when the agent is incapable of interpersonal or social relationships of a certain kind due to irrationality, one can consider how to affect an agent's behaviour and entertain a reactive attitude when the agent is rational and responsible. One can realize what caused an agent to act in a certain way, and so consider how to prevent or promote similar actions in the future, while at the same time realizing how the attitude revealed by the action relates to one's expectations of the agent, and so experiencing reactive attitudes
on that basis. An objective attitude may be evident after a reactive one subsides, but that does not mean that it was not present when the reactive one was as well. In fact, the reverse may occur: after one has decided how to interact with an agent to promote or deter certain actions or affect certain attitudes, one may still have reactive attitudes concerning the agent. There is no conceptual incompatibility between reactive and objective attitudes, and if Strawson's claim is the psychological one that considerations of how to affect an agent's behaviour are incompatible with an emotional response to the agent's attitudes as revealed by action, this seems to be empirically false. One may consider the causes of an act and still have feelings of gratitude, resentment, pride or shame, for what is at stake is an evaluation of character, and how a person came to have a character of a certain type if irrelevant to assessing what types of character that person has, though considering how a person came to have a certain character would not be incompatible with an assessment of character and an emotional response to that assessment based on like or dislike.

Reactive attitudes are therefore inherent in certain types of relationships, but the expression of those attitudes in word or deed is not. The latter is a pragmatic question, based on what likely effects such expression will have on the agent, and these pragmatic considerations are not incompatible with the experience of reactive attitudes. To act toward agents solely on the basis of rational and pragmatic considerations of how best to control or affect their behaviour
does not, and could not, rule out that important element of life which consists of emotional responses to the attitudes of agents, and so it would be foolish to act on the basis of those responses (what Strawson calls reactive attitudes) when such action would be harmful for fear that not doing so would eliminate that important element of the inner life which consists in solely having those emotional responses. One can then care about agents and about oneself and still behave rationally and objectively toward them. (In some cases, of course, expression of one's feelings may be both appropriate and necessary, but the appropriateness of the feeling does not entail the appropriateness of its expression. If one does not respond with suitable affection to kind acts one may eliminate the possibility of a desired relationship with an agent because one's lack of favourable response may be interpreted as a lack of incapacity for feeling; if one does not respond with anger to certain sorts of hostility, that may rule out certain relationships with others who infer from one's apparent lack of indignation a failure to take others seriously or a lack of concern about oneself.) Whether to express one's feelings in praise or blame, then, is a pragmatic question, for such expression follows neither from responsibility of the agent (which is a necessary condition of the appropriateness of the feeling), nor from the appropriateness of the feeling. While they are elements of praise and blame, then reactive attitudes do not illuminate when praise and blame should be expressed and reveal more about the expectations inherent in interpersonal and social relationships.
than about responsibility.

In this section we have seen how responsibility is compatible with determinism, and how it relates to practices such as praise and blame. Determinism rules out the legitimacy of some attitudes of praise and blame that hold the agent totally and ultimately responsible for what the agent is, but it does not rule out evaluations of acts coupled with an assessment of responsibility. Nor does it rule out praise and blame as means of controlling behaviour. Lastly, it was shown how the emotional component of some attitudes of praise and blame, known as reactive attitudes, do not follow from responsibility, but reveal more about the nature of relationships between persons.
PART THREE

AUTHENTICITY AND FREEDOM

1. Introduction

The first part of this essay showed how the concept of responsibility illuminates the structure of action, but that it is the structure of action itself that provides a basis for both the concept and the practice of responsibility. It was found that responsibility increases with agency, and agency is the degree to which the effect of the action is under the control of the agent. Actions are then a kind of event typified by the fact that they are produced by a system of causes within the agent, and are thus events over which an agent has control. In order for an agent to have control over an event, the causes of that event must include some deliberate or purposive exercise of power by the agent, and thus for an event caused by an agent to be an action the system of causes within the agent that produces the event must include an intention. Realization of intentions constitutes a greater degree of agency, as it entails greater conscious control over the action, and it was pointed out how different adverbs of excuse point out different ways in which an agent may fail to control an action and not realize his or her intention. Control over one's intentions, by formulating them in accordance with one's goals and knowledge (and formulating the latter
two in accordance with the evidence available to one), is what constitutes rationality, which is a condition of full control over an action, or of full agency. The thesis which emerges from this analysis is that greater agency entails greater responsibility; and hence is what freedom is in the sphere of individual action. Freedom of action is then the degree to which the individual has control over an action and its effects.

This concept of freedom was defended in the second part of this essay against the determinist or incompatibilist position, which equates freedom with causal indeterminancy, or "contra-causal" freedom. It was shown that since freedom is agency, it is not an action's being caused that makes it free or unfree, but the nature of those causes. To the degree to which an agent has control over an action, and so to the degree to which the causes of an action are within an agent and subject to that agent's conscious control, the action is free. How it came to be that the agent was such that he or she had control over the action is irrelevant to any considerations of whether or not the action was free. The first point of the second section explored the limitations determinism does place on our concept of freedom and specifically attempted to show how some notions of praise and blame are based on a concept of contra-causal freedom which is untenable. The relationship of standards of behaviour and of praise and blame to responsibility was then investigated.

This final section deals briefly with what the existentialists have dubbed "authenticity." The reason for including a
discussion of authenticity here is that it is an extension of the concept of freedom that has been developed so far. By assuming responsibility for one's values, and hence for one's goals and objectives, one achieves a greater degree of control over one's actions and life, and hence a greater degree of freedom. We will argue that freedom is a matter of degree (and not, as Sartre says, absolute) and that authenticity is the highest degree of personal freedom. While agreeing with the existentialist that freedom is founded on the existence of a consciousness that can conceive of its situation being other than it is, and so act to bring about a desired state of affairs, this is only a necessary condition of freedom, for freedom in the full sense is not merely the power to intend, but the power to control these intentions and to realize them. Full freedom is an agent's control over an action, a situation or a life, and the freedom which is the nature of human consciousness (the freedom to intend) is not a sufficient condition of that full freedom.

2. Degrees of freedom

The reason that freedom is a matter of degree should be clear given an analysis of agency. For the degree to which an action is the successful realization of the intention of the agent, the agent has control over it, and so is free. On the level of intentions, to the degree to which the intention is rational, it is subject to the control of the agent, and to that degree it is free. Freedom is then the extent to which actions are realizations of intentions and the extent
to which these intentions are rational, so that an action free in that it is rational may fall short of full freedom if the intention is not realized, and an action that is the realization of an intention may fall short of freedom if the intention is not rational. Full freedom thus coincides with full agency.

It is thus a degree of control and of freedom to realize intentions, and another degree of control and freedom to make these intentions subject to one's goals and knowledge. To control one's goals and values, and hence the intentions and actions on which they are based, is then to realize an even greater degree of freedom, at the same time as it is to assume a greater degree of responsibility. Authenticity together with full agency is then the highest degree of freedom.

3. Value and choice

The existentialists argue that values are not given or a priori as discernable facts in the world, but are choices made as a solution to the problem of what to do, and thus are a basis of action and inseparable from the concept of action. As how to act is a question which faces each individual, the values chosen through action are chosen by the individual, and so the individual, the values chosen through action are chosen by the individual, and so it is the individual who is responsible for them. Traditional world view and value systems have no guarantee of being right; they do not have the status of mere descriptions of properties of objects or situations, as they
contain an imperative or prescriptive quality: a "Thou shalt" or a "Thou shalt not." As they are not factual claims, value systems are unverifiable, i.e., they are not logically deducible from any objective state of affairs or from a value-free description of a state of affairs. As a value is not a fact, it is not discoverable by reasoned analysis or by observation, and so reason cannot demonstrate to us which values to choose, as values are not inherent in or deducible from things. (Even self-preservation can take second place to other values, and so utilitarian calculuses based on any notion of self-interest that is not circular, in that it is not defined simply as the satisfaction of the individual's values, for which it is supposed to serve as a foundation, are as open to question as any other values, and also reflect some value choice not deducible from any fact, such as a valuation of one's own life over all else.) Valuation is then an individual act of choice, not a discovery of fact, and so one is responsible for one's own values.

Value systems take on the appearance of a fact when they serve as a means of escaping the necessity to choose, since to accept an already formulated value system is to accept the choice of others without realizing that one has chosen to accept that choice; and so it is to take those values as facts, rather than choices. To accept a value system while denying to oneself that one has chosen to accept it hides the evaluative act contained in accepting that system (or its representatives, such as society, the Church, the Party, etc.) as authoritative. A value is not worthy of being chosen simply because many, or a whole society, hold it (since 50 million French can be and
have been wrong), even if one could choose to value society as one's authority on values (thus choosing to abdicate choosing for oneself. As Kierkegaard points out, to do so would be foolish because one is in danger of hiding one's choice from oneself and of regarding society's value choices not only as authoritative, but as factual, in which case one would be abdicating one's freedom and responsibility to choose, but un成功地, as one would have that freedom and responsibility even if one hides that fact from oneself). For example, to decide that the good is that which the gods desire and so to decide that if something is desired by the gods, it is good, is not to make an empirical discovery, but to choose a standard for all one's valuations (eg., that which the gods desire is good). To valuate is then, as Nietzsche put it, to legislate for oneself without recourse to basing that decision on a value inherent in an object or state of affairs, since objects in themselves are value-less, and receive value only through their valuation by some person.

It is of no avail to seek to escape the necessity of choosing values by referring to linguistic conventions. "The fact that other human beings have evaluated certain situations in certain ways, and have succeeded in building those evaluations into the language available for describing those situations has no automatic normative implications," as Olafson points out in Principles and Persons. Words used in moral discourse (eg., "promise," "lie," etc.) have a descriptive element that refers to behaviour which is separable from their evaluative component and hence does not logically entail the evaluative
or prescriptive component. Words containing both normative and descriptive elements do refer to practices, but they do not establish the justifiability or value of those practices. To say that a person is a necromancer once implied that that person should be regarded as evil and so destroyed, but it does so no longer (save for those who choose to accept a value system that includes the value that necromancy is noxious or evil, which is the missing premise of the practical syllogism that "if a person is a necromancer, destroy that person").

The individual is then faced with a choice that has no support in the nature of things or the opinions of others, and that unfounded choice is the foundation of that individual's values. The individual must choose in order to act, since action is inseparable from the postulation of values (in that all actions must have a goal). Kierkegaard points out that to choose an action is to choose the way of life implicit in that action, and to endorse that choice as right and hence as absolute (though one could change one's mind and renounce one's choice later, which would require a new choice and so a new endorsement of a way of life). To act is thus to posit a value which through universalizability over situations (though not over persons, as one cannot choose for another) becomes a standard of action. In order to exist, one must act, and in order to act, one must make value choices. Valuation is then inseparable from human existence. The individual is faced with the necessity of choosing values and total responsibility for that choice, in that it derives from nothing save that individual.
One cannot base one's choices on one's past or on a concept of human nature, since in choosing values one chooses how to live, and in choosing how to live one chooses what to become. As Heidegger and Sartre point out, human existence has no fixed nature (in the sense that it has no pre-determined telos, even if it is definable biologically as a species), and so entails no natural values. Rather, each individual chooses his or her nature in choosing how to live.

4. Sartrean freedom

This doctrine of self-creation relies on a theory of the freedom of human consciousness which has in its most famous explanation in Sartre's chapter on "Being and Doing" in Being and Nothingness. Sartre points out that the individual subjective consciousness separates itself from that which it takes as objects for consciousness: for any object of which consciousness is aware, consciousness is aware that it is not that object, as it is the thing aware of that object. In other words, for any thing perceived there must be a thing perceiving (consciousness) which is not that thing perceived. Consciousness is thus a pure activity discoverable in contradistinctions to the objects towards which it directs itself and which it sees it is not.

Consciousness is then not to be confused with the ego or the self, which is a construction by consciousness of its past actions and states and which consciousness transcends or separates itself from by its positing of the ego as an object for consciousness. Consciousness cannot then take its
past as a reason or justification of any course of action, which is directed toward the future, in that consciousness is not defined by its past since it transcends it. Moreover, the ego can only be defined by consciousness in terms of the ends toward which consciousness projects itself. Thus, it is the values one chooses which determine how one regards oneself, and not the other way around, in the sense that one's view of oneself cannot be determined unless one is using some criteria, such as a value or ideal, so that one sees oneself in terms of that value or ideal which one seeks to realize. It does not follow, as Sartre seems to imply, that causal factors do not determine how one regards oneself, and hence how one chooses values, nor does it follow that the ego which is consciousness' construction of its past actions and states is the same as the ego which is the actual totality of the self (rather than the perceived totality of the self). The point is that, since consciousness is separate from the ego (by which we mean the ego consciousness postulates as an object for itself), it cannot take the ego as a justification or reason for choosing a certain value, for it is in choosing a value that consciousness defines or confers meaning upon the ego. Thus, as Kierkegaard says, one does not act morally because one is moral, but one is moral because one acts morally; in acting morally, one chooses oneself as a moral being. Or, as Sartre says of Genet, one does not steal because one is a thief; one is a thief in virtue of the fact that one steals, and so to attempt to justify one's behaviour on the basis of one's being a thief is simply to say that one steals because one
steals, which of course is no justification of one's behaviour at all. To say one steals because one is a thief masks one's choice to continue to remain a thief by continuing to steal, and that choice cannot be justified by the mere fact that one has stolen in the past. One is then left with the necessity of defining oneself through action, and so with choosing values which may serve as a basis for an answer to the question of how to act.

Consciousness is thus not the self which it takes as an object for itself, and it cannot take itself as an object for itself, because as soon as consciousness posits itself as an object, it is separated from itself as an object for consciousness in the same way as it is separated from any other object for consciousness, and so consciousness as subjectivity transcends itself as object. Despite this, consciousness strives to define itself (i.e., turn itself into an object for itself) by projecting itself toward the realization of its possibilities (and define what it has been in terms of those possibilities). The future toward which consciousness directs itself is given shape by the way consciousness views its past and thus confers upon its past a meaning. (Past stages are seen as stages toward the stage toward which one is projecting oneself and which one has chosen, and so are valued as means or obstacles to that project.) The future toward which consciousness directs itself is conditioned by its present in that its future goal is seen to be lacking in the present situation.

One is therefore continually surpassing one's situation, including the self that one has been. This is what is meant
by the dictum that "existence preceded essence." One first of all exists and then defines oneself. One is not already defined by some pre-existent human nature, for though there are elements of one's situation that are due to one's being human, one is not defined by one's situation, but is free to transcend it by endowing it with meaning (by projecting oneself toward future goals). Situations in themselves do not motivate action, but our appreciation of them (i.e., the way we interpret them) does. A goal is a state of affairs not yet existing, and so it is not an existing state of affairs but that which is seen as lacking in it that produces an intention to act. The situation cannot determine consciousness to perceive a lack of it, but the lack is seen when a project or value is chosen to be realized. Intentions are produced by an apprehension of something that is desired, and what is desired is a non-existent state of affairs, as even a continuation of the status quo is a future possibility not existent in the present, and so seen as lacking. Thus, insofar as actions require intentions, and intentions require that something not yet achieved be achieved or striven for, we can agree with Sartre that action arises from a consideration of a situation as a lack of a desired possible. In that actions require intentions, they result from choice. (Although those intentions need not be realized in order for an event to be an action, as failure of an action is not simply the non-realization of an intention, but it can be the effecting of unchosen or unthought of states of affairs, as our analysis of action revealed. Thus, for example, Sartre says a man who drops a cigarette and causes a
fire does not act, whereas we could say he acted, as he meant to drop the cigarette or he meant to smoke, but the result of his action (the fire) was inadvertent and so unintended. Yet, there was an intention (to smoke, to drop the cigarette) and so there was action.) Even though to perceive an existing situation as lacking is to perceive the existing possibility of a different situation, and that possibility is grounded in the present situation, the situation is seen as revealing a possibility, rather than a mere fact, only by the negation of the present situation by the positing of a future, as yet non-existent state of affairs as a goal, that is, by a choice. The way in which one's situation is viewed is thus determined by the ends toward which consciousness projects itself. Action is directed toward a future which does not yet exist and which is posited by consciousness in the form of possibilities, which are products of consciousness in relation to its past and to its situation, and not of the past or situation itself. One's situation only limits the choice of one's ends, it does not determine that choice (in that no imperative is deducible from any situation). What one takes for a "cause" of an action, by which Sartre means a reason for acting, is not a result of the situation itself but of the end one has chosen, and thus a "cause" is an objective situation perceived in view of a presupposed end. Similarly, a "motive" is a subjective state of affairs (particularly, one's emotions and desires) interpreted in light of a chosen goal. Thus, what Sartre calls "causes" and "motives," by which he means facts about oneself in the latter case and facts about one's situation in the form-
er that are taken as reasons for acting, are determined (in the sense of being defined) by the ends one chooses, and so do not determine those ends. Thus, motives, causes and ends are posited at once, and each is defined in relation to the others. Hence, facts about the situation or oneself do not cause an intention in that it is only in the presence of an intention that those facts become reasons for acting. Any fact taken as a reason for action at one time loses its force in the absence of an intention that confers upon it its character as "cause" or "motive," and so a fact taken as a reason for acting at one time need not be taken as a reason at another, and will only be taken as a reason if one again posits the intention which makes it so. The facts of one's situation thus do not determine one's intentions, for they are seen as determinants or reasons for action only in light of an intention.

There is a great deal of confusion over Sartre's theory of freedom, as it can be interpreted in two ways. Taken as a theory concerning the justification of actions it seems to be most correct, and this is the way in which we have taken it. The explanation of action with which Sartre is dealing refers to the agent's reasons for acting and the justification of action on the basis of those reasons. One's situation is given, but how one views it (e.g., what aspects of it are taken as reasons for acting) is a result of a choice, which is not determined in the sense that a situation cannot produce justifications or reasons for one course of action rather than another, even if the choice is causally determined (by physical or psychological factors). Reasons for action are due to one's goals,
and for them one is always responsible, as goals are the result of a choice of value which is not determined by objective facts in that no values exist \textit{a priori} or in the nature of things. Determinism is therefore simply irrelevant. Even if my action is caused, my intention is the product of my own choice and not of any objective fact about myself or my situation. Objective facts acquire value as causes or motives only in view of an end which has already been posited, and which is grounded in a value chosen without justification, as at bottom the value on which other values rest cannot be justified in terms of any other value, but must simply be chosen. In this Sartre seems correct.

Yet if one takes Sartre as defending an indeterminist theory of freedom, his position is utterly confused. The confusion is between the reason or justification for an action and its actual cause. What Sartre means in saying that consciousness cannot be determined in its choices is that no objective facts can persuade or compel it to embark on a certain course of action, as facts in themselves are value-less and action requires the positing of value. But it does not follow that choices of value arise \textit{ex nihilo} in the sense that they are uncaused. To argue that is to neglect the fact that consciousness is caused (physically and psychologically) by things other than itself, even if it transcends those things in the sense that it can take them as objects for itself. This is biology: no interpretive metaphysics can save Sartre here. There is also the problem that although I am free to give meaning to my situation in view of my goals or projects, the
situation still has an effect on my choice, not as a reason but as an actual (rather than as a perceived) cause of action. For example, if a past is being fled as unwanted, does the person who flees that past enjoy the same freedom as one who does not have to escape it? Certainly there are alternative ways of fleeing the past, but one's choices will have to be made with reference to that particular past, and so are circumscribed. To be sure, one can deny one's past by dying, but that in no way mitigates the causal influence of the past upon one's present choice. The same can be said about other aspects of my situation, which Sartre groups together under the concept of "facticity." Sartre recognizes that facticity limits choice (as it limits alternative course of action), but in denying that it does not causally determine choice he confuses logical freedom with psychological freedom, and falsely deduces from the fact that no situation justifies an action the much stronger claim that no situation causes an action.

5. Authenticity as freedom

It is Sartre's theory of freedom as a consideration of the justification of actions which forms a basis for authenticity as a value, and so his failure to refute determinism is of no consequence. By recognizing one's power to choose ends and values, one achieves a degree of freedom and autonomy lacking in most, even if ultimately one's ends are determined by physical and psychological causes (since the way one views one's situation may be causally determined by psychological and social factors beyond one's control), for one is
simply more aware of what one is doing and why, in which case one is at liberty to reassess one's goals and values. To step outside of the value system one has been presented with and realize that values are a matter of one's own choosing is to realize a greater degree of freedom as it is to exercise a greater degree of control over one's intentions. To be aware of one's freedom to choose values is to be in a better position to realize that freedom, and to be aware of the necessity of choice and of one's own responsibility for choosing is to be made aware of that freedom. In that way, one can make sense of a passage in Sartre's *Situations*:

> We were never more free than during the German occupation ... the choice that each of us made was face to face with death, because it could always have been expressed in these terms: "Rather death than ..." 41

To act in a fully free manner one must act authentically, that is, with the recognition that one has chosen and is responsible for the value upon which one has acted, as it has no basis or source outside of one's free choice. In doing so, one chooses a way of life and a world (in the sense that one contributes to the creation of oneself and the world through action and in endorsing the value upon which the action is based one chooses oneself and one's world, and not in the sense that one's choice is subject to a Kantian categorical imperative, as so many have wrongly taken Sartre to be saying). By realizing that one creates one's life and world through choosing values and acting upon them, one endows one's life and world with meaning. To flee from this realization is not only cowardice, but an abdi-
cation of freedom. To decide to not consciously decide is to renounce control over one's intentions, and so narrows one's possibilities by making them a matter of chance, such as social circumstance, prevailing mores or other causal factors. On the other hand, to choose in awareness of one's freedom and responsibility for that choice to leave open the permanent possibility of choosing anew, of revaluing or renouncing old values, and thus one's possibilities are widened. 42

Since authenticity is the recognition of one's responsibility for one's values, it is also the recognition of one's responsibility for oneself. One is responsible for oneself only insofar as what one is is the sum of what one does, and what one does is the result of values for which one is responsible. One is not responsible for oneself in the sense that one is an *en causa sui*, for one cannot pre-exist one's own existence. But one can choose how to be by choosing how to act. Even though one is always making oneself and is never made or complete as long as one exists (as one continually transcends one's past), one can be aware of choosing what to become, and so exercise control over what one will become. One is able to, as Nietzsche put it, fashion one's life as a work of art, in conformity with a value or ideal one has chosen as admirable. To realize one's freedom to choose one's values and hence one's self and thus to choose on the basis of that realization is authenticity, and this being so, authenticity is a realized ideal even if complete self-definition is not (due to consciousness' continual transcendence of itself). 43
To be in control of one's values is to be in a better position to realize values, for in choosing values in awareness of the fact that I have chosen them I am aware of what my values are and of the fact that they are within my control since they result from my choice. Authenticity, which is the conscious exercise of control over one's intentions by being aware of choosing the values upon which those intentions are based, thus gives one a measure of control over one's life in that one is able to consciously choose how to live, and so allows for self-realization, which is not just the inevitable fulfillment of one's possibilities, but the fulfillment of those possibilities one has chosen to fulfill. To not be aware of one's values is to be unable to intentionally realize them (though one could realize them by accident), and to not be aware of choosing one's values is to sacrifice one's control over the life one lives and the person one becomes. Freedom is a basic value because it is necessary in order to intentionally realize any other value, and authenticity as a form of freedom is a basic value in that it is necessary in order for the individual to be in control of his or her values, rather than subject to them.

Authenticity has been questioned as an ideal. Grene says that Sartre sees authentic individual existence as an end in itself, whereas, she claims, authenticity is a by-product of acts, not an end or purpose. But we have seen that freedom is a basic value in that it is necessary to the realization of any other value, and as authenticity is a form of freedom is a basic, not secondary. Moreover, the distinction Grene
makes between what is chosen and how it is chosen needs to be argued for: the whole corpus of Kierkegaard's work is an argument to the effect that what matters in order for life to have value or meaning is not that values are chosen, but how they are chosen, and choosing them in awareness of one's act of choosing (and with complete concern for one's choice) allows one to place a meaning or significance on one's existence. For Existentialists, the quality of life is its most important aspect and therefore most to be sought, so it is the real end of life; the ends by which it is reached (the values chosen authentically) by authentically directing action toward those goals are secondary. Granted, if one is to be authentic, one must have values which one choose authentically. But surely no one will deny this, and so this cannot stand as a criticism of authenticity as an ideal. If it is a value to be in control of one's life, then authenticity, which is to be in conscious control of one's values (upon which one's life is based) must be a value.

The popular misapprehension of authenticity as an ideal that calls for capricious behaviour is dispelled when one realizes that actions are based on chosen values and principles, and so are governed by them. Acts are arbitrary only insofar as value choices are arbitrary, but the fact is that all value choices must rest on some basic choice that is not justifiable by reference to any other choice or by any objective fact, and so are arbitrary, but to be aware of choosing one's values in fact makes them less arbitrary, as they are placed under one's control, rather than resulting from the chance workings of determinism alone. To be authentic is then the
opposite of acting capriciously, and freedom is thus the control of one's values and hence of one's actions, not allowing one's life and actions to run out of control. Since freedom is control over one's life and actions, it is impossible to be free and not be responsible, and in realizing one's responsibility for one's values one places greater, not less, control over one's life and actions.

The idea that values are chosen has also come under criticism, but as Olafson points out, this too is due to a misperception of the Existentialists' position. To say an action is founded on a choice is not to say that a mental event occurred which one may or may not remember, but that there is no possible description of an intentional act which does not refer to choice. Choice can refer to a mental event rather than to an action, as for example with "She chose a theatre career, but changed her mind and took up law," but it can also refer to action with no implication of a mental event, as with for example "Cato chose death over dishonor." It is clear that the Existentialists have the latter sort of choice in mind, not the former, and as actions cannot occur without intentions, choice is a necessary element of action.

6. Conclusion

As we have seen, however, as a theory of what freedom is, authenticity is incomplete. Control over one's values is but an element of freedom, an element which, to be sure, is essential to freedom in its fullest form. But a theory of freedom which does not take into account the freedom which constitutes realizing
one's choices, as well the freedom which consists in consciously making those choices, is faulty. Moreover, Sartre seems to think that it is mere ability to make choices that is freedom, and while we can agree that such an ability is the basis of freedom, it is the actual taking of control, the actual authentic making of choices which constitutes freedom, for the mere fact of choosing and being able to choose is not the same as consciously choosing, and thus being in control of what one chooses. Choosing values while suppressing the knowledge that one is choosing is hardly freedom in the way that awareness of choice is freedom, as in the former case choice is left to chance, even if it does originate from the individual, while in the latter it is under the control of the individual in the only real sense of control, and that is under conscious control. Thus, awareness of choosing is freedom, choosing is not, and choice is only an element of freedom, since full freedom also requires the realization of choice.

Authenticity is thus the conscious control of one's values through the realization that one chooses one's values and is therefore responsible for them. As such, authenticity is freedom, as it is control over one's actions, as actions are based on values, and hence of one's life. Freedom is a value as it is essential to the true realization of any other value, for unless one is in control of whether or not a value is realized, one cannot realize it intentionally, and authenticity as freedom is a value in that it allows for self-realization by placing the individual's values, and hence the individual's life and actions, within the individual's own conscious control. Never-
theless, control over one's actions through control over the values on which they are based is only partial control and partial freedom; full control and full freedom require that the values over which one exercises control by consciously choosing them be rationally realized.
CONCLUSION

We have seen then a coincidence between responsibility and freedom. For an agent to be responsible for an action, the action must have been under the control of the agent. Control over the action consists in the realization of intentions, which is control over the effects of the action, and the rationality of those intentions, which is control over the intentions. Full responsibility thus requires full control of the action. Yet there is another, fuller sense of control over actions that is achieved when the agent consciously chooses values with the realization that they exist as values for the agent due to the agent's choice and hence the agent is fully responsible for them. This greater degree of control is a greater degree of freedom, but it is also a greater degree of responsibility, as it is the assumption of responsibility not only for one's intentions and their realization, but of the values on which the intentions are based, and so is the maximum degree of responsibility an agent can have for an action. It is, in fact, the assumption of responsibility for one's life plan, as it is taking responsibility for one's goals. Thus, it is responsibility not only for each action individually, but for one's pattern of actions. It is then, when it is combined with the rational realization of intentions, the highest degree of responsibility and freedom.
NOTES


3. Ibid.

4. cf. Part Three, pp.133-140


12. Coval and Smith, "The Concept of Action."

13. Ibid.


17. Ibid, pp. 118-121.


20. Ibid, p. 27.


22. Ibid, p. 35.

23. cf. p. 64 above.

24. The degree to which one is responsible for one's character is discussed in Part Three, particularly on pages


30. Hare, R.M., _The Language of Morals_, (London: Oxford University Press, 1952)


32. Ibid, pp. 111-114.

33. Ibid, p. 132.


38. Sanborn, p. 112.


42. Grimsley, pp. 102-103.


47. Olafson, p. 165.


49. Olson, p. 108.

BIBLIOGRAPHY


