AN ANALYSIS OF STRAWSON'S DEFENCE OF MORAL PRACTICES AGAINST THE CLAIMS OF DETERMINISM

by

ALEXANDER HOLTBY BOSTON

B.A.(Hon), The University of Saskatchewan, 1993

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

The Faculty of Graduate Studies Department of Philosophy

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

September 1995

© Alexander Holtby Boston, 1995
In presenting this thesis in partial fulfilment of the requirements for an advanced degree at the University of British Columbia, I agree that the Library shall make it freely available for reference and study. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by the head of my department or by his or her representatives. It is understood that copying or publication of this thesis for financial gain shall not be allowed without my written permission.

Department of Philosophy

The University of British Columbia
Vancouver, Canada

Date Aug. 2, 1995
I plan to examine and defend Strawson's argument in "Freedom and Resentment" against various criticisms from Downie, Russell, Ayer, Watson, and Nagel. Strawson holds that moral responsibility and determinism are compatible, exploring people's feelings and attitudes to make the case that people neither can nor should give up reactive attitudes, whether or not determinism is true. I will use Wolf and Dennett to supplement Strawson's argument, and examine Bennett's paper to see whether certain modifications of Strawson's view are necessary.

Using character improvement as its foundation, I will offer an account of responsibility compatible with Strawson's argument and thus with determinism. I will rely on Dennett's concept of the ability to act and Taylor's idea of reflective identification with one's actions to construct my account.
# TABLE OF CONTENTS

Abstract ii

Table of Contents iii

Acknowledgement v

Chapter I: Introduction 1

Chapter II: Strawson's Argument
   i) The Basic Dispute 5
   ii) Personal Reactive Attitudes 8
   iii) The Moral Reactive Attitudes 14
   iv) The Threat of Science 19

Chapter III: Initial Objections
   i) R.S. Downie's Objections 25
   ii) Paul Russell's Objections 28

Chapter IV: Replies to the Initial Objections
   i) A Defence of the Voluntary Objective Attitude 35
   ii) Strawson's Rationalistic and Naturalistic Strategies 43
   iii) A Token-Naturalistic Defence of Self-Reactive Attitudes 45
   iv) A Token-Naturalistic Defence of the Other Reactive Attitudes 50
   v) A Defence of our Rationality 56

Chapter V: Ayer's Critique
   i) Predictability and Determinism 66
   ii) Science and the Objective Attitude 70
   iii) The Applicability of Desert 73

Chapter VI: Watson's Partial Definition of Responsibility
   i) The Global Excuses Taken Individually 78
   ii) The Unfortunate Childhood 85
   iii) Living Counterexamples: Einstein, Gandhi, and King 88

Chapter VII: Bennett's Attempt to Define Responsibility
   i) Minor Additions to Strawson's Theory 99
   ii) The Difficulty of Defining the Reactive Attitudes 101
   iii) The Objective Attitude and Teleological Inquiry 104
   iv) Interpersonal Relations and the Reactive Attitudes 107

Chapter VIII: On "Moral Luck"
iv

i) Variable Consequences and Moral Tests 113
ii) Character, Attitudes and
    Contra-Causal Freedom 119
iii) Christian-Kantian Thought 124

Chapter IX: Responsibility
i) Reactive Attitudes and Reflection 130
ii) The Basis of our Excuses 132
iii) Why we want Responsibility 138
iv) Minor Objections 139
v) On Freedom 141

Chapter X: Conclusion 149

Bibliography 152
I would like to thank Dr. Earl Winkler, whose criticism and encouragement helped me to try harder. In addition, I am grateful to Dr. Catherine Talmage for her assistance. I am also indebted to Iris Winston, Stewart Boston and David W. New, whose editing skills helped to make this thesis readable.

I gave a talk formed from part of chapter IV at a Philosophy Department Graduate Student Colloquium at the University of British Columbia on Dec. 5, 1994, and am grateful to those who attended, particularly Paul Russell, Adam Constabaris, Richard Johns, Michael Fleming and Alan Richardson, for their helpful comments.
Chapter I: Introduction

For centuries, philosophers have pondered the problem of free will. Most people do not worry about this problem. They simply hold themselves and others responsible for actions that according to common sense are freely chosen.

Yet some people believe that free will is an illusion, and that, since moral responsibility depends upon free will, it too is an illusion. They assert that for any supposedly free action there are causes, which have other causes, and so on. This chain of causation extends backward beyond the birth of the agent, placing these early causes well beyond her control. Because of their belief in such a causal chain, these people are determinists. They are also incompatibilists, since they believe that this causal chain makes free will and moral responsibility impossible.

Other people think that free will exists. They hold that human beings can make choices which are not wholly determined by any temporally extended causal chain. These people are indeterminists because they deny unbroken causal continuity in the field of human action. Yet they do not consider human choices to be random. These people are also incompatibilists because they too believe that causal continuity makes morality impossible. Because of their belief in some form of contra-causal free will they are libertarians.

A third group opposes the incompatibilism of both the previous groups. They are compatibilists, and they believe that
free will, or at least moral responsibility, is compatible with causal explanation of human action. Usually, compatibilists argue for the social usefulness of holding people morally responsible. Incompatibilists find this utilitarian account of responsibility to be unsatisfying.

P.F. Strawson defends a compatibilist position in his paper "Freedom and Resentment."¹ His approach, however, is different from that of most other compatibilists. Strawson refrains from any direct discussion of the thesis of causal determinism. Instead, he begins with an account of certain emotional attitudes, called reactive attitudes, which people are prone to take in reaction to the actions or attitudes of others. He then investigates whether the legitimacy of these reactive attitudes, which express feelings associated with holding people responsible, are threatened in any way by the possible truth of determinism. He argues that their legitimacy remains intact and then extends this argument to include the specifically moral reactive attitudes.

I believe that Strawson's account provides a solution to the problem of moral responsibility in a determined world. This problem is the major component of the more general issue of free will and determinism.

I will elaborate Strawson's approach in chapter II. In chapters III-V I will raise and address various criticisms of Strawson's compatibilist view. These criticisms usually come in two forms. The first form of criticism is that Strawson fails to
show adequately that reactive attitudes would persist if belief in determinism became widespread. The second is that reactive attitudes are irrational if determinism is true. I believe that Strawson’s position can be defended against both criticisms.

In chapters VI-VIII I will examine possible consequences of Strawsonian compatibilism. I will also investigate various attempts to define responsibility, isolating a few elements which should be excluded from its definition. In chapter IX, I will give my own definition of responsibility, and show how it fits within Strawson’s view. Like Strawson’s account of reactive attitudes, my definition of responsibility is compatible with determinism.

Strawson’s article is worth reading for anyone who believes in determinism and yet thinks moral responsibility is something more than an illusory device to control people’s behaviour within society. Strawson helps to alleviate concerns which compatibilists might have about the apparent irrationality of holding such seemingly disparate views.
Chapter II: Strawson's Argument

i) The Basic Dispute

P. F. Strawson starts by confessing ignorance about what determinism is.¹ I offer the following definition for the reader. Without going into explicit detail, one description of determinism is that every event, including every human action, is caused by other events, which in turn are caused by others, and so on. The chain of causation extends backwards until the beginning of time (if there is one) or indefinitely (if there is not).

Strawson does not require an explicit definition of determinism. Since one can portray determinism roughly, Strawson notes that, if a coherent and true thesis of determinism exists, then, according to its definition of determinism, all human behaviour is determined.²

Many people claim to know what determinism is, and Strawson examines their principle arguments. He briefly mentions the "genuine moral sceptics," who hold that, whether or not determinism is true, concepts such as moral blame, moral guilt and moral responsibility are confused and thus inapplicable. But Strawson mainly concerns himself with a debate between the "optimists," who hold that these concepts and their associated practices, such as punishment and moral condemnation, have a place in a determined world, and the "pessimists," who hold that determinism renders moral concepts inapplicable, and the practices associated with them unjustified.³

The labels "optimist" and "pessimist" are Strawson's.
Scientists are adept at finding causes for events, making some form of determinism plausible. Yet, most people would regard the loss of our moral concepts and practices as undesirable. Strawson therefore calls those who believe determinism to be incompatible with applicable moral concepts pessimistic, and those who hold the opposite view optimistic.

The better to analyze the debate between these views, Strawson renders the optimist’s viewpoint formally: "(1) the facts as we know them do not show determinism to be false; (2) the facts as we know them supply an adequate basis for the concepts and practices which the pessimist feels to be imperilled by the possibility of determinism’s truth." The optimist supposes that moral concepts and practices are valid in a wholly determined world. She notes how powerfully punishment, moral condemnation and approval can regulate the behaviour of an individual in socially desirable ways. Moritz Schlick, for example, supports this view: "Punishment is an educative measure, and as such is a means to the formation of motives, which are in part to prevent the wrongdoer from repeating the act (reformation) and in part to prevent others from committing a similar act (imitation)." Thus, the optimist holds that moral concepts are valid because they help to control people’s behaviour for the benefit of society.

But Strawson’s pessimist says that the optimist’s account is insufficient to ground the morality of these practices. For punishment to be just and not merely useful, for condemnation to be moral and not merely beneficial, people must have moral guilt.
This requires moral responsibility, which requires freedom, which in turn requires that determinism be false.

The optimist responds that freedom and determinism are still compatible, if one defines freedom negatively. The optimist notes that negative freedom requires the absence of certain external and internal conditions, such as compulsion by others, physical restraints, and insanity. It also requires the absence of various forms of ignorance or accidents with respect to the agent. For example, if Jane imagines a piece of paper to be some rubbish and tosses out a friend’s winning lottery ticket, she is excused from responsibility because of her ignorance. In the presence of such ignorance, punishment or condemnation cannot be expected to have any desirable effect on the regulation of Jane’s behaviour. To be held responsible, she must be free of all of the above conditions.

But to the pessimist, this form of freedom is not enough. In addition to the absence of certain excusing conditions, freedom requires a genuine identification of the will with the act. The optimist allows for this, and says that a person can have reasons for acting, decide to act because of these reasons, and do just as she wants to do. She thus identifies her will with the act. This is part of the facts as we know them, and still allows for the possibility of determinism. The optimist can even allow the possibility that the agent could have performed a different action - provided she had so willed. But given all causal antecedents exactly as they were - which include everything about her
background, make-up and particular situation - the optimist would hold that the agent could not have willed other than she did.\textsuperscript{11}

The pessimist replies that, for those cases in which negative freedom is present, the only justification the optimist gives for moral practices is their efficacy in regulating behaviour. This seems to most people to be an insufficient basis for moral practices.\textsuperscript{12} There is more to morality than the manipulation of others for social purposes. The pessimist requires that the agent somehow could have acted differently in her actual situation, exactly as she was and exactly as it was.

So, as the debate now stands, we can side either with the optimist, who says that most people simply must revise their understanding of the basis of moral practices, or with the pessimist. If the latter, we must either, like the sceptics, give up any attempts to justify moral practice, or else try to formulate a kind of contra-causal freedom that requires the falsity of determinism. This sort of freedom is difficult to define coherently, especially since one wants it to be not just a description of random events but a basis for morality.\textsuperscript{13}

\textit{ii) Personal Reactive Attitudes}

All of this is familiar. What Strawson does is not. He gives the optimist another view of the foundation of moral practices - a view consistent with determinism, but satisfying the pessimist's demand to justify moral concepts as more than devices of social regulation.
Yet, Strawson does not discuss moral practices immediately. He first speaks of "the very great importance that we attach to the attitudes and intentions towards us of other human beings, and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about these attitudes and intentions." To use a variation of Strawson's example, although the physical pain is similar whether Albert steps on Susan's hand accidentally, or with malicious intent, Susan will feel "a kind and degree of resentment" in the latter case absent from the former. Even when no specific harm is done, a person will typically feel resentment towards another's display of hostility, as when someone is rude or insulting.

We demand some form of goodwill or consideration in our interactions with others. The form this demand takes will vary with the relationship, as will the kind and degree of our feelings towards those who show or fail to show goodwill.

Reactive attitudes are reactions toward the attitudes and intentions of others or ourselves. To understand reactive attitudes better, Strawson focuses on resentment. Resentment is a personal reactive attitude, that is, a reaction to the attitudes of others towards oneself. Because it is personal, it is not in the same category as moral indignation or guilt. Yet, Strawson feels that by examining this cousin of the moral reactive attitudes, we will make better initial progress, since the field of personal reactive attitudes, unlike that of moral reactive attitudes, is uncluttered with moral theories.
Strawson examines the situations in which we feel resentment, and the special considerations which limit or cancel this feeling. He divides these considerations into two broad categories. First, there are such excuses as accident or ignorance, which still permit us to view the antagonist fully as an agent. These considerations remove or limit feelings of resentment towards an agent; we see the agent as wholly or partially excused for the injury she caused. Yet although these excuses may restrain our resentment towards the agent in this case, we still may feel reactive attitudes towards the agent in general. Thus, such considerations

"...do not suggest that the agent is in any way an inappropriate object of that kind of demand for goodwill or regard which is reflected in our ordinary reactive attitudes. They suggest instead that the fact of injury was not in this case incompatible with that demand's being fulfilled, that the fact of injury was quite consistent with the agent's attitude and intentions being just what we demand they should be."\(^{18}\)

On the other hand, Strawson's second category contains excuses that cause us to view the agent in a new light. Strawson subdivides these excuses into two groups: those which excuse short-term conditions, and those which excuse long-term, even lifelong, conditions. An example of the first sort is a man under an excessive amount of stress, of whom one could say, "He was not himself" if he should commit an injury while stressed. Our resentment against the injury he causes is limited. We usually do not have to encounter the person when he is under so much stress, and thus we "shall not feel resentment against the man he
is for the action done by the man he is not."\textsuperscript{19} This excuse will be examined in chapter VI.

The second sort of circumstance is important for Strawson’s argument. The excuses found within this group disqualify a person from being viewed as a responsible agent. They protect certain people, such as small children and the insane, from blame. When this sort of excuse is made for someone, we are asked to suspend indefinitely all reactive attitudes towards her, including resentment. Instead, we revert to what Strawson calls the "objective attitude."\textsuperscript{20}

For instance, when Mary takes the objective attitude towards Jane, Jane is seen not as a complete person, but to some extent as an object. Although Mary might feel emotions such as pity or fear towards Jane, Mary will feel neither resentment nor any other reactive attitude. Mary could consider Jane to be an "object of social policy," or "a subject for...treatment," but could not be involved in a fully human relationship with her.\textsuperscript{21} The objective attitude

\textquote[...cannot include resentment, gratitude, forgiveness, anger, or the sort of love which two adults can sometimes be said to feel reciprocally, for each other. If your attitude towards someone is wholly objective, then though you may fight him, you cannot quarrel with him, and though you may talk to him, even negotiate with him, you cannot reason with him. You can at most pretend to quarrel, or to reason, with him.\textsuperscript{22}]

Strawson notes that, in addition to cases in which we naturally take the objective attitude, we can voluntarily take it towards normal people. We can use his resource, as Strawson puts
it, to relieve the pressures of personal involvement, or as an aid to government or institutional policy. When we so treat responsible agents, however, we are never entirely successful, nor can we retain the attitude for long.\textsuperscript{23}

Since we can choose to take either the objective or the reactive attitudes towards normal people, can or should a belief in determinism cause us to give up our personal reactive attitudes? This is the crucial question that Strawson answers next. He looks first at the cases in which reactive attitudes are limited or eliminated, starting with the agent whose capacity for normal interpersonal relations is not in general limited, someone who acted through ignorance, accident, or other exceptional circumstance. Strawson notes that a belief in determinism does not explain the limitation of resentment in this sort of case. For, obviously, determinism does not entail that all the injuries that agents generate fall into this category. People could still bear malice. Total lack of malevolence "would not be a consequence of the reign of universal determinism, but of the reign of universal goodwill."\textsuperscript{24}

The second case of taking the objective attitude applies when the other is seen not to be a responsible agent at all. For instance, she may be a child, or have a severe psychological abnormality. Strawson notes that it cannot be a result of determinism that we are all constantly under abnormal stresses, that we are all psychologically abnormal to the point of incapacity, or that we are all children.\textsuperscript{25}
Finally, Strawson looks at the cases in which we voluntarily adopt the objective attitude towards normal people. If determinism is true, can or should we take the objective attitude towards others at all times? In answer to whether we could, Strawson notes that although it is not logically inconsistent to suppose this, it is practically inconceivable that we be able to give up our reactive attitudes permanently. Even a widespread belief in determinism would not be enough to overcome our strong commitment to the ordinary interpersonal relationships that make human society human.\textsuperscript{26}

In addition, Strawson notes that when we do take the objective attitude towards someone, whether abnormal or normal, it is not because of some belief in determinism. We always have other reasons, as listed above. Determinism, which affects all people if it affects any, does not provide a reason for taking the objective attitude towards particular people.\textsuperscript{27}

After saying that human beings almost certainly cannot take the objective attitude towards all people all the time, Strawson tries to answer the question of whether, if determinism holds, we rationally should take the objective attitude towards everyone. At this point, Strawson posits first that our commitment to interpersonal attitudes is part of the general framework of human life. Although we can judge particular cases within it, we cannot assess this entire framework as rational or irrational.\textsuperscript{28} Consequently, whereas we can decide on the rationality of resenting a particular person for a particular action, we cannot
judge if the general disposition to feel resentment and other reactive attitudes is rational or irrational. This disposition is something we are simply given with human life as we know it.

Second, Strawson states that even if we could step outside ourselves and judge whether or not interpersonal reactive attitudes as a whole are rational, such a judgement would have to be based, not on the truth or falsity of determinism, but on instrumental reasoning concerning the enrichment or impoverishment of human life.²⁹

iii) The Moral Reactive Attitudes

Strawson shows that our personal reactive attitudes, such as resentment, are independent of considerations of determinism and that we would not necessarily be more rational to allow its truth to affect them. He then turns to the moral reactive attitudes. These are often felt on behalf of a third party but can be felt on one's own behalf, considering oneself as a member of the moral community. For instance, although John may resent Bill's punching him in the nose, he would not resent Bill's punching a total stranger. Rather, he would be morally indignant with Bill for punching the stranger. Nevertheless, John could simultaneously feel resentment for Bill's punching him and moral indignation that Bill would punch anyone. Strawson calls the category of reactive attitudes that includes moral indignation the vicarious analogues of personal reactive attitudes.³⁰ I will call these the vicarious reactive attitudes for the sake of convenience.
Strawson also mentions a third type of reactive attitude, called the self-reactive attitudes. Guilt, shame, and some forms of pride fall into this category, for they are all attitudes associated with the demands one makes of oneself towards others (and occasionally towards oneself). Although self-reactive attitudes are also moral reactive attitudes, Strawson focuses upon the vicarious reactive attitudes.

The three types of reactive attitudes - personal, vicarious and self-reactive - are, if not essentially connected, at least humanly linked. Strawson has difficulty imagining a person having one or two types of these attitudes without the other(s). Theoretically there could be moral solipsists who have no vicarious reactive attitudes, or some strange form of saints with no personal reactive attitudes. Perhaps there are some psychopaths with only personal reactive attitudes, and only some of those. Yet these individuals are "far below or far above the level of our common humanity" and thus out of Strawson's scope.

Next, Strawson looks at the excuses we make for people in connection with the vicarious reactive attitudes. Like the personal reactive attitudes, their vicarious analogues involve a demand for some kind of regard from the person towards whom they are directed, although this regard is typically demanded on behalf of another instead of oneself. Similarly, the two types of excuses that wholly or partially inhibit resentment and other personal reactive attitudes also wholly or partially inhibit the vicarious reactive attitudes, and in the same way. One minor
difference Strawson notes is that, since we are not as personally involved, when we excuse a person because she is not fully an agent, we put more emphasis on her lacking moral responsibility.  

Furthermore, we can voluntarily adopt the objective attitude in moral cases, just as we can in the personal cases. Yet, the motives to do so are fewer, since in moral cases, as opposed to personal cases, we lack the strains of personal involvement from which we need to seek relief.

Strawson draws further parallels. As with the personal reactive attitudes, when we suspend our moral reactive attitudes we never do so because of a belief that the action in question is determined. We cannot conclude from determinism that everyone is always ignorant of the consequences of her actions, or always acts accidentally, or is psychologically abnormal, or is a child, or is incapable of guilt.

In addition, Strawson notes that since the personal and moral reactive attitudes are very much interconnected in human society, the failure of personal reactive attitudes to decay despite belief in determinism suggests that moral reactive attitudes would also survive. That is, we are probably as incapable of living without moral reactive attitudes as we are personal reactive attitudes, regardless of belief in determinism.

As to whether it is more rational to adopt the objective attitude in moral cases, Strawson submits that, just as with personal reactive attitudes, moral reactive attitudes as a whole cannot be brought up for rational review. Furthermore, if we
could somehow choose whether or not to have them, our criteria used would again be pragmatic and instrumental and have nothing to do with determinism.38

To show further that the objective attitude is unrelated to determinism, Strawson uses examples of a child and a psychiatric patient. In these cases, the parent or psychiatrist takes not a wholly but a partially objective attitude, which shades into normal reactive attitudes as the child grows or the patient heals. The partially objective attitudes in these two cases have nothing to do with determinism.39 If determinism does or should have an effect on our taking the objective attitude, it would require us to take it, not partially or towards a few people, but fully and towards everyone.

Finally, Strawson returns to the dispute between the optimist and the pessimist. The pessimist is unhappy with the optimist’s view, which holds that morality makes sense under determinism solely because of the desirable social consequences of regulating behaviour. Strawson sees a reason for this unhappiness. In keeping with this justification of morality, the optimist seems to suggest an objectivity of attitude towards all people at all times. People are regarded as subjects for control and treatment. But the optimist, by taking (or rather, by implicating) the objective attitude even towards actions that meet the conditions of negative freedom, ignores the moral dimension itself, namely the moral emotions and attitudes.40 "A thoroughgoing objectivity of attitude, excluding as it does the moral reactive attitudes,
excludes at the same time essential elements in the concepts of moral condemnation and moral responsibility."  

Moreover, when we partially withdraw our goodwill, which we usually do when we feel negative reactive attitudes, we do not exclude the offender from our moral community, as we would if we took the objective attitude. On the contrary, we continue to view her as a member of the moral community, only as a violator of that community's rules. This presents a basis for punishing offenders. Although we may hope that such punishment has a positive effect in preventing future transgressions, it is the withdrawal of goodwill entailed by our reactive attitudes, and not the utility of punishment, that is essential to understanding the morality of punishment.

Thus, to apply moral concepts and justify moral practices, we must examine the "web of attitudes and feelings which form an essential part of the moral life as we know it, and which are quite opposed to objectivity of attitude." The optimist should modify her reasons for permitting morality in a determined world. In turn, the pessimist should concede that both the personal and the moral reactive attitudes are compatible with determinism, and, therefore, that morality is compatible with determinism. Strawson's argument bridges the gap between the optimist and the pessimist.

Even the moral sceptic, whom Strawson excludes from his main argument, can benefit from it. The sceptic sees that the "one-eyed utilitarianism" of the optimist is inadequate to explain what
people mean by morality, and yet that the pessimist’s descriptions of counter-causal freedom seem unintelligible, and difficult to relate to morality. Thus, to the sceptic, moral concepts are inherently confused. But Strawson’s moral reactive attitudes provide an intelligible basis for moral concepts and practices that covers all that people mean by morality. Thus, the sceptic may comfortably give up scepticism as readily as the pessimist may give up pessimism.  

iv) The Threat of Science

After presenting his argument, Strawson answers some objections to it from the scientific quarter — or, at least, from philosophers with pessimistic interpretations of scientific facts. First, different human societies have different expressions of human reactive attitudes. A joke in one culture might be a horrible insult in another. Thus some might object that it is ethnocentric to say that all societies have reactive attitudes. Still, there is no human society that completely lacks all forms of reactive attitudes. Such a society might not even be intelligible to us. Thus, Strawson can insist that reactive attitudes are basic to human society.

Second, it is perhaps true that some manifestations of our reactive attitudes are symptoms of our unconscious unfulfilled desires, or of other unconscious processes, and so are untrustworthy as moral guides. Strawson answers that our reactive attitudes’ having causes does not invalidate their
appropriateness.\(^47\) To point out that diamonds are merely compressed coal does not make them less beautiful to look at. Furthermore, I would add that it seems unlikely that all our reactive attitudes have unconscious rather than conscious causes. Therefore, many of our reactive attitudes are untouched by this objection.

The third objection arises from how scientists study humanity. They can detach themselves from their subjects, and the scientific method has led to the modification of some aspects of our reactive attitudes. Scientists discovered, for example, that diseases have a biological origin and not a divine one, and are, therefore, not divine punishments for evil deeds. Some believe that such discoveries will eventually destroy our reactive attitudes. Still, philosophers must take all facts into account. The scientific observer is also human, and thus is not completely separate from her reactive attitudes. Strawson thinks it extremely unlikely that "our progressively greater understanding of certain aspects of ourselves will lead to the total disappearance of those aspects."\(^48\) He does allow that it is not entirely inconceivable, however.

To conclude, Strawson agrees with the optimist that moral concepts are applicable and moral practices are justifiable, even if determinism holds. Nevertheless, although these moral practices help to regulate the behaviour of offenders, and although if certain moral practices are ineffective, we may modify those practices, the optimist should keep in mind that these
practices are expressions of our moral attitudes, indeed of our humanity itself, and not mere regulatory devices. "Our practices do not merely exploit our natures, they express them". It is only by understanding this that we can understand why our moral practices are effective. When we look at the system of reactive attitudes from which these moral concepts and practices originate, we can understand why our moral concepts have meaning beyond pragmatic usefulness, and why our moral practices are both justified and effective.


4. I will ignore possible random events on the quantum level.


8. Strawson, "Freedom and Resentment", 73. Strawson actually phrases it as "a genuinely free identification of the will with the act", but I do not wish to use the word "freedom" in the definition of freedom and omit the word.


13. The first of these views is known as soft determinism, and is held by Schlick, Hobbes, Hume and others. The second of these views is known as hard determinism, and is held by Paul Edwards, Schopenhauer, Freud and others. The last of these views is known as libertarianism, and is held by C.A. Campbell, Roderick Chisholm, and others. "Hard and Soft Determinism" by Paul Edwards, New York University, has more detailed comparison of the two types of determinism. His paper was apparently given at the 1957 American Philosophical Association meeting. Unfortunately, no further information about where the paper was printed, if anywhere, is currently available.


21. Strawson, "Freedom and Resentment", 79. Of course, one might be related to the person in question without having the sort of relationship that Strawson describes. For instance, one might have a daughter whose mental deterioration is such that she cannot be considered a responsible agent, and one cannot be involved in a certain type of relationship with her, since that type of relationship would require one to have reactive attitudes towards her. I believe that Strawson calls this type of relationship a human relationship because only humans are capable of it, although not all humans are.


27. Strawson, "Freedom and Resentment", 83.


29. Strawson, "Freedom and Resentment", 84. Strawson adds in the footnote on this page that in a sense we would be more rational creatures if we were guided solely by the objective attitude, but that "it would not necessarily be rational to choose to be more purely rational than we are". But in "P. F. Strawson Replies", Philosophical Subjects, Zak van Straaten, ed., (Oxford: Clarendon Press, 1980), 261, he makes clear that he meant rational in two different ways in that sentence. That is, it would not necessarily be in our rational (pragmatic) interests to choose to be more purely rational (emotionless, or ratiocinative) than we are.


31. Strawson, "Freedom and Resentment", 86.

32. Strawson, "Freedom and Resentment", 89. Although Strawson does not include self-reactive attitudes within moral reactive attitudes in his original article (See Strawson, "Freedom and Resentment", 93), he does do so in a later article ("P.F. Strawson Replies",
Solipsism is the view that the only thing one can be sure of is one's own existence. Thus a solipsist holds that the rest of the world is illusory. A moral solipsist, therefore, holds that she alone is entitled to resent how others treat her and that she alone is entitled to feel guilt at how she treats others. She feels no moral indignation at how others treat others, since if the action does not specifically involve her in some way, it is not real to her.

34. Strawson, "Freedom and Resentment", 86.
38. Strawson, "Freedom and Resentment", 89.
42. Strawson, "Freedom and Resentment", 93.
44. Strawson, "Freedom and Resentment", 94.
46. Strawson, "Freedom and Resentment", 95-6. One could put a number of psychopaths on an otherwise unpopulated island and observe what they did, and this would be a society of a sort. I doubt that such a society would be recognizably human, or that it would last for long.
47. Strawson, "Freedom and Resentment", 96.
Chapter III: Initial Objections

Strawson does not concern himself with an exact definition of determinism, and argues that whether determinism is true or false is not relevant to morality. Strawson examines the excuses that cause us to disengage our reactive attitudes in certain cases or towards certain people. He shows that an all-encompassing thesis such as determinism gives us no reason to take the objective attitude towards everyone. This is because our entire emotional framework cannot come up for rational review. Thus, it cannot be judged irrational. Furthermore, even if we had such a reason, we would be unable to give up our reactive attitudes completely. Thus we neither can nor should give them up in toto, even if determinism is true.

Strawson faces heavy criticism from R.S. Downie and Paul Russell. He dismisses too quickly, according to Downie, the importance of the metaphysical question of whether or not determinism is true. According to Russell, Strawson fails to argue successfully that, if determinism is true, we cannot or should not overcome our tendency to hold reactive attitudes.

i) R.S. Downie's Objections

Downie looks at the excuses that inhibit reactive attitudes, specifically the excuses that cause us to take the objective attitude, and believes that they have a common element. "It is tempting to say that what promotes the objective attitudes is the belief that the operation of these factors renders the behaviour
open to sufficient explanation in causal terms.\textsuperscript{1} Strawson talks of resorting to the objective attitude when one considers a person to be a subject for treatment and control, and this seems relevant to Downie's proposal: control and treatment are not useful unless they have a causal effect.\textsuperscript{2}

Since to think of something in causal terms is by some definitions of determinism to place it within a determined system, Downie believes that Strawson unfairly undermines the pessimist's position. When Strawson says that we can voluntarily take the objective attitude towards normal people, he implies that we can explain normal behaviour in causal terms, just as we can that of abnormal people. He further implies, therefore, that all people are determined. Thus Strawson no longer merely argues that determinism might be true; rather, he already assumes determinism's truth in a premise of his compatibilist argument.

Downie objects to this sneaking in of determinism. He thinks that it implies that the reactive attitudes are reducible to the objective attitude. That is, one can always ignore the agency of a person, and the reasons or intentions she gives for her actions, and instead take the objective attitude to determine the "real" cause. Furthermore, Downie holds that the objective and reactive attitudes are not simultaneously justifiable. If someone's actions are explicable in causal terms, it is appropriate to take the objective attitude. It is therefore inappropriate to take the reactive attitudes. If we can take the objective attitude towards anyone at all, then everyone's actions
are explainable in causal terms; so, it is inappropriate to take the reactive attitudes towards anyone. If Downie is correct, then determinism, if true, affects the appropriateness, or rationality, of taking the reactive attitudes.3

What is appropriate is not always possible, however, and Downie allows that Strawson may be right to say that it is impossible for human beings to do without reactive attitudes. Downie even agrees that, if we had the choice, it might not be desirable to give up the reactive attitudes. Nevertheless, Downie holds that despite what it may be rational to do, the metaphysical point of what is rational to believe if determinism is true is important.4 If Strawson assumes determinism and relies on this assumption in his compatibilist argument, then the metaphysical point is not argued, but, rather, settled arbitrarily.5

In defence of the libertarian viewpoint, Downie gives an alternative description of Strawson's cases of voluntarily taking the objective attitude. Downie holds that Strawson uses the word "objective" ambiguously. When he uses it to describe involuntary adoption of the objective attitude, it seems to mean the severing of our human relationships with that attitude's object. Yet, in voluntary cases, we do not take the objective attitude, but rather a cool reactive attitude. This attitude is impartial and detached, but regards its object as a fully responsible human being; thus, it is a reactive attitude, albeit at the cool end of the emotional scale.6 A judge deciding a court case takes this attitude; to take the objective attitude would be morally
objectionable.\textsuperscript{7} Thus, Downie explains our occasional voluntary
detachment from people without assuming the truth of determinism,
as he believes Strawson does.

Strawson apparently fails to answer the pessimist. Downie’s
account of voluntary detachment does not imply determinism, as
Strawson’s apparently does. Therefore, Downie holds that Strawson
has insufficiently argued for the rationality of believing that
we are morally responsible beings, if determinism holds.
Nevertheless, Downie allows that, from a pragmatic point of view,
it is rational for us to act as morally responsible beings,
whether determinism is true or not.

\textbf{ii) Paul Russell’s Objections}

Paul Russell yields much less to Strawson. He believes that
Strawson fails to establish that it is rational, even if
determinism is true, to act as if we are morally responsible
beings. He also believes that Strawson fails to establish the
impossibility of abandoning the reactive attitudes. Russell
divides Strawson’s arguments into the rationalistic and the
naturalistic. The rationalistic argument is meant to answer the
pessimist’s complaint that determinism’s truth is incompatible
with morality, and, thus, gives us reason to reject the reactive
attitudes. The naturalistic argument is meant to show that, even
if we had a reason to reject the reactive attitudes, we are unable
to do so.

Russell believes that Strawson’s arguments interfere with one
another. Since the naturalistic argument holds that no amount of evidence for our moral incapacity can enable us to reject our reactive attitudes, the rationalistic argument, which asserts that we are not morally incapacitated, seems superfluous. This is a minor point, however, and Russell finds larger problems within each of the two arguments.

Strawson declares in his naturalistic argument that we are unable to eliminate our reactive attitudes. Russell believes that this is because these attitudes, like fear, are emotional reactions. Whereas instances of these emotional reactions may be inappropriate, anyone who claims that the disposition to have an emotion is itself inappropriate attempts to judge an emotional framework externally, an attempt at which people are doomed to fail. People can only make internal judgements of the times when a certain emotion is unsuitable. Therefore, we are psychologically stuck with emotions such as fear and the reactive attitudes.

Russell believes that this argument is an instance of type-naturalism, which asserts that we cannot give up our capacity to feel emotions. The argument counters a type-pessimist, who holds that the general belief in determinism and its consequences would cause us to lose our reactive attitudes. A pessimist of another sort, however, accepts the type-naturalist's claims. She agrees that we cannot give up our susceptibility to resentment and gratitude any more than we can give up our susceptibility to fear. Nonetheless, she notes that we are able to judge and give up an
emotion in particular instances when we realize that it is inappropriate. She concludes that we are also able to give up every instance of a reactive attitude towards every person, should we discover that all human beings are parts of a determined world. This is the argument of the token-pessimist. If we became immune (or better still, realized that we had always been immune) to all dangers, we would cease to feel instances of fear, although retaining the capacity for it. So, too, we could give up all instances of resentment, gratitude, or any other reactive attitude if we discovered determinism to be true. We would retain the capacity for such emotions, but they would never be used.\textsuperscript{9}

Russell believes that there are counterarguments to the token-pessimist. The first points out that in numerous instances, emotions are warranted. This is true of fear. Yet, it is not true of the reactive attitudes, if determinism's truth morally incapacitates everyone. The other counterargument, the token-naturalist's, is that no argument of the pessimist could cause us to give up instances of feeling any of our emotions or, at least, the emotions that the token-naturalist wishes to defend. This argument does not seem persuasive enough to defend the necessity of our feeling fear. Russell maintains that he can conceive a situation in which we would feel no more fear. By analogy, Russell argues that if the token-pessimist is correct, then a general belief in determinism would cause us to stop feeling any reactive attitudes; everyone would count as a case of moral incapacity, which requires resort to the objective attitude.\textsuperscript{10}
Thus, type-naturalism does not address the token-pessimist’s concerns, and token-naturalism is not convincing.

Russell next criticizes Strawson’s rationalistic argument. Russell believes that this argument, if sound, adequately defends our retaining the reactive attitudes. The argument purports to show that it is a mistake to hold that determinism entails that everyone is morally incapacitated. The argument insists that those cases in which we involuntarily take the objective attitude have nothing to do with the implications of determinism. Strawson states that we must resort to the objective attitude when "...the agent is seen as excluded from ordinary adult relationships by deep-rooted psychological abnormality - or simply by being a child. But it cannot be a consequence of any thesis which is not itself self-contradictory that abnormality is the universal condition."11

Russell examines those cases in which, because we do not consider a person to be an agent, we excuse her of responsibility. Strawson’s argument is that psychological abnormality cannot by definition be true of everyone; yet if determinism’s truth were the reason for this abnormality, then the abnormal would be universal and we would have a logical contradiction. Russell points out that it is not because psychological abnormality is abnormal that we reject the psychologically abnormal as agents. Rather, it is because the abnormality is some form of incapacity, which merely happens to be abnormal. It is conceivable that most or all people could become morally incapacitated, such as from a
widespread mental illness; this would make moral incapacity normal. Strawson's contention that the notion of widespread abnormality is incoherent does not address the possibility of widespread incapacity. The token-pessimist argues that, if determinism holds, moral incapacity is universal. Thus, if Strawson means that it is incoherent to say that an abnormality is normal, then the token-pessimist will agree, but the incapacity she is talking about, if it exists, is no abnormality.

If, on the other hand, Strawson is merely stating that we do not in common practice draw the line between agency and nonagency such that it will exclude everyone from agency, then the token-pessimist will reply that this is because most people believe in free will. They either reject determinism or fail to see that its consequences preclude free will's possibility. The token-pessimist believes that the line would be redrawn to exclude all people from agency if enough people both believed in determinism and realized its consequences.

The token-pessimist seems to be unchallenged. Strawson's rationalistic argument is unconvincing, and his naturalistic argument does not apply to her. Determinism might entail that all people are incapacitated, and people might give up their reactive attitudes because of a belief in determinism.

Furthermore, Russell believes that Strawson's argument that human beings cannot rationally suppress their emotions even when they realize the facts is disturbing. Russell thinks that Strawson's naturalistic strategy
"casts doubt upon our ability or capacity to curb or control our emotional life according to the dictates of reason...[The strategy] invites us to accept or reconcile ourselves to reactive attitudes...even in circumstances when we have reason to repudiate them."\(^{13}\)

Strawson holds that if we were able to judge our reactive attitudes as a whole, we would consider whether they enriched or impoverished our lives, as opposed to worrying over determinism.\(^{14}\) Russell opposes the optimist's belief that, if determinism entails complete moral incapacity, it would be more enriching to hold reactive attitudes towards people than to live in accordance with the facts.\(^{15}\) Russell holds that it is possible to have good reasons to give up the reactive attitudes. A token-pessimist values living in accordance with the facts, and she believes that determinism entails moral incapacity for all. Since Strawson's arguments do not convince her, she argues that, if determinism holds, we can and rationally should give up our moral reactive attitudes altogether.

Downie and Russell seem to show that Strawson's argument has failed in its compatibilist aims. The schism between the pessimist and the optimist has opened up once more. Fortunately, it can be repaired.
1. Downie, R.S., "Objective and Reactive Attitudes" in Analysis vol.27 no.2 (new series no.116), Dec., 1966, 34.

2. Downie, 33-4.

3. Downie, 34-5.

4. Downie, 36.

5. Downie, 36.

6. Downie, 37.

7. Downie, 38.


Chapter IV: Replies to the Initial Objections

i) A Defence of the Voluntary Objective Attitude

I first address the concerns of R.S. Downie, who believes that Strawson has implicitly assumed determinism. According to Downie, the ability to explain behaviour in causal terms is an essential prerequisite for the objective attitude. If it is possible to take the objective attitude voluntarily towards anyone at all, everyone's behaviour becomes causally explicable.

We must look closely at this line of reasoning. When we try to control or treat those toward whom we take the objective attitude, we make predictions about how they will react, basing our guesses on our knowledge of a person, of people relevantly like them, or even of people in general. Sometimes the objects of our attitude will foil our guesses, acting unpredictably and inexplicably. We are by no means as accurate in our predictions of human action as in our calculations of the position of an orbiting comet.

More importantly, determinism does not have to be true in order for predictions to be useful. If determinism is false, and people enjoy contra-causal freedom, it is still possible to predict their actions roughly. Libertarianism does not entail that people's actions are totally unpredictable. For instance, if someone chooses chocolate ice cream over vanilla one hundred times in a row, one may be able to guess fairly accurately that she will choose similarly the next time.

This guess could be wrong. Nevertheless, we do not become
more accurate predictors if determinism holds. Even should determinism be true, the causes of our actions are diverse and complex. The best we can do is make educated guesses, or, for large groups, statistical analyses. Too many factors can affect a person's actions to allow completely accurate predictions, given the limitations of the human mind and, indeed, of any computer made of matter. Nevertheless, we make reasonably accurate predictions whether we live in a deterministic or indeterministic world.

When Strawson introduces the voluntary objective attitude, he is not assuming determinism. People are, at best, only partially successful in explaining another's behaviour. And if determinism is false, we can still guess someone's likely actions in a given situation.

To avoid the supposedly deterministic implications of Strawson's objective attitude, Downie introduces the cool reactive attitude. With this attitude a judge can control his passions and decide a divorce case in a disinterested fashion, while still considering the people involved to be free and responsible.

I agree that people can take a cool reactive attitude. I think that Strawson would allow the introduction of cooler and hotter reactive attitudes as a gauge of emotional involvement. Nevertheless, in addition to cases of the cool reactive sort, cases that fit the voluntary objective model remain. For instance, a politician can choose to forget about treating people as people for a moment in order to consider whether a new law
will control crime, or increase her popularity. She thus considers how the law will motivate and influence people. She contemplates people's possible actions. She does not concern herself with their feelings except insofar as they will affect her constituents' actions. Another example is someone who uses pick-up lines in bars in order to get dates. Here, the only effect considered is how to influence others with certain stimuli to attain desired ends. When taking the objective attitude, we consider a person's deliberations and decisions to act to be unimportant, except instrumentally; we focus on how to make her behave as we desire. For prediction and control purposes, it is irrelevant whether the responses are generated by people or by objects. This indicates the voluntary objective attitude rather than the cool reactive attitude. Therefore, while Downie's cool reactive attitude adds an interesting new perspective on reactive attitudes, it does not replace voluntary cases of the objective attitude.

Since the objective attitude is not tied to determinism, the voluntary objective attitude now belies Downie's view that taking the objective attitude renders the reactive attitudes inappropriate. One can switch between the voluntary objective attitude and the reactive attitudes towards normal people, depending on one's focus. The two types of attitude have different purposes, but to adopt one does not make it wrong to take the other.

Finally, Downie believes that, no matter what is possible for
people, or even what it is pragmatically rational for people to do, what we rationally believe concerning metaphysical fact is still important. Determinism's possible truth is more relevant than Strawson thinks.

We can examine these issues in two ways. Either determinism should rationally have some effect on why we consider people free and responsible beings, or it should not. Strawson believes the latter. He holds that the framework of our reactive attitudes is not up for external rational review. If no reason exists for determinism to affect how we consider ourselves, then the question of determinism becomes purely intellectual. On this interpretation, our moral concepts and practices, and our beliefs about ourselves as morally responsible beings, are not affected, nor should they be.

Let us assume, however, that determinism necessitates that we are other than free and responsible beings. In this case, whatever utilitarian reasons we possess to retain our reactive attitudes, we are not living in accordance with the metaphysical facts. Because of this, we seem to have a reason to change our attitudes towards ourselves. Thus, people are in some sense not entirely rational if they believe in determinism and yet maintain reactive attitudes towards themselves. This seems to be Downie's concern.

Susan Wolf addresses this concern in "The Importance of Free Will". Wolf agrees that if determinism both is true and provides a metaphysical reason to give up our reactive attitudes, failing
to do so will leave unrealized the value of living in accordance with the facts. Wolf holds, however, that even if determinism holds and implies that we are unfree and unresponsible, we have "no reason at all to regard ourselves as unfree, unresponsible beings" (emphasis hers). Thus, we leave no value unrealized in retaining our reactive attitudes. Nothing is lost.

Wolf considers a variety of cases. First, she looks at a drug addict who cannot stop taking a drug. To avoid circularity, we assume in this case that most people are free and responsible beings, whether this requires indeterminism or not. The addict may be content in his addiction, or he may try everything in his power to stop taking the drug. He is not free to resist taking the drug. Both we and the addict himself, however, consider him to be free and responsible in his attitudes. Perhaps he approves of his drug taking and wishes to take responsibility for it. On the other hand, he may be an unwilling addict who feels compelled to take the drug and, therefore, accepts no responsibility for taking it. Whatever attitude he takes, we would consider him to be free and responsible at least in making the claim that he does or does not wish to take the drug and in accepting or rejecting responsibility for taking it. This is what it means to say that he asserts his freedom as an attitude-taker by adopting an attitude towards his drug-taking, though he is unfree as a drug-taker.  

On the other hand, if determinism is true, and if this entails that we are all unfree and unresponsible, then Wolf
postulates that we should consider ourselves to be like robots. We would lack control over not only each of our actions but also each of our apparent choices. If so, the pessimist argues, we should, to be rational, take the objective attitude towards everyone, including ourselves.

But doing this will not help us realize the value of living in accordance with metaphysical fact. This is because "in taking any attitude toward ourselves, including the attitude that we are not free or responsible beings, we would be asserting ourselves as free and responsible beings" (emphasis hers). In taking an attitude, a person affirms that she is free and responsible, at least with respect to attitude-taking, even as she declares that she is not. In making a judgment, she declares that her judgement matters, and that she is responsible for the judgement. In effect, she separates off a part of herself, which remains free and responsible, and then disclaims responsibility for the rest. By saying, "I am not responsible for any of my actions," she presupposes the existence of an "I" who is divorced from her actions, yet responsible for denying responsibility for those actions. Determinism may entail that nobody is at all free and responsible. Yet, in implying that we are not responsible beings, it also implies that we are not responsible even insofar as we are attitude-takers. Therefore, we are no more rational to take one sort of attitude towards ourselves than another. Both contradict the idea that we are completely unfree and unresponsible.

We have a third choice besides taking the objective attitude
towards everyone and holding the reactive attitudes: we might cease to have any attitudes at all. We would make no judgements about anyone's responsibility or lack thereof. Thus, we would not affirm our freedom or responsibility by making judgements. Of course, for us to do this is as impossible as it is to take the objective attitude at all times, but the issue we must decide here is whether it is rational for us to give up all our attitudes, reactive and objective.

If we did surrender all our attitudes, we would have to change how we perceive ourselves dramatically. Obviously, we would have to give up notions of responsibility and desert. We would also have to give up ideas of what ought to be, since this would lead to what ought to be done (by oneself or others), which would in turn lead to the notion of responsibility. We would stop thinking of some things as being better than others in any normal sense. Values would be no more than desires. "We would lose our ideals, our senses of self, and...our status as persons."8

Even given these ill consequences, do we have any reason to live in accordance with the facts? Wolf thinks not, even after she takes into account that people often wish to live, not in a world of comforting illusions, but in the "Real World."9

We seem to have only two reasons to live in accordance with the facts. First, we possess a desire to "get things right." Yet, if we give up taking any attitude towards anything, we also give up such value-concepts as right and wrong. By doing this, we leave ourselves no right way to live. We remain with our value
of getting things right unrealized, because the abandonment of our attitudes entails the abandonment of this value.\textsuperscript{10}

Our second reason to live in accordance with metaphysical fact is that we "want ourselves to matter in the right sort of way, to make the right sort of difference to the world and the beings who do matter and to whom we might matter."\textsuperscript{11} But if determinism is true and entails that we are not free and responsible beings, then its effects are universal. By abandoning all attitudes towards everyone, we eliminate all values. The result of this is that no beings would matter, and so, we could not "make the right sort of difference" to any "beings who do matter and to whom we might matter." Thus, we cannot realize the goal of mattering by giving up all attitudes. By these arguments Wolf shows that we have no reason to give up all of our attitudes.\textsuperscript{12}

Wolf counters Downie's argument that metaphysical facts concerning determinism can provide any reason to give up our reactive attitudes. She shows that determinism rationally should not affect our conception of ourselves as free and responsible beings. The truth or falsity of determinism once more remains a matter of mere intellectual interest; it has no effect upon the rationality of moral practices and concepts. This answers the last of Downie's objections. Thus, although we can adopt the voluntary objective attitude towards others without presupposing determinism, we would not be more rational to give up the reactive attitudes, even if determinism is true.
ii) Strawson's Rationalistic and Naturalistic Strategies

Russell's objections are harder to answer. To begin with, there is an apparent conflict between Strawson's rationalistic and naturalistic strategies. The rationalistic strategy argues that determinism fails to entail moral incapacity. It shows that determinism is wholly unrelated to our reasons for giving up our reactive attitudes. The naturalistic strategy, on the other hand, argues that it is psychologically impossible for us to give up all reactive attitudes "even if some general truth were a theoretical ground for it". Since Strawson's rationalistic strategy argues that determinism provides no reason for us to give up our reactive attitudes, does he undercut his own naturalistic strategy? Does the rationalistic strategy imply that we can give up our reactive attitudes if determinism provides a theoretical ground to do so?

I think not. First, the rationalistic strategy maintains only that determinism has nothing to do with the excuses that cause us to give up our reactive attitudes involuntarily. It says nothing about our voluntarily giving up these attitudes if we feel that determinism warrants it. By showing that, when we excuse people of responsibility for an action, we do not use determinism as a basis for the excuse, Strawson simply disconnects determinism from our actual practice and its grounds. If he fails to do this, then his naturalistic strategy becomes more difficult. It would be harder for him to argue that people are incapable of giving up the reactive attitudes altogether if an excuse that currently
causes us to give up our reactive attitudes had determinism as its justification. It would still be possible to accept the naturalistic strategy. This strategy argues against our giving up the reactive attitudes completely, rather than only when we face one of the common excuses people use. But if such an excuse is motivated by determinism, then, given determinism's universality, we seem arbitrarily to excuse some people but not others.

It makes the naturalistic strategy more compelling if determinism is independent of any such excuses. This restricts any force determinism possibly has to cases in which we voluntarily give up the reactive attitudes or to some excuse never before used. We can hold the objective attitude towards normal people for a short time only. Strawson's naturalistic argument therefore becomes more believable. We can more easily believe it to be psychologically inconceivable due to a belief in determinism that we can permanently give up our reactive attitudes, since our voluntary attempts to give them up are of such short duration, and all known excuses that cause us to give them up are unrelated to determinism. The naturalistic strategy does not require the rationalistic strategy to succeed, but the latter helps to make the more controversial naturalistic strategy more credible. Therefore, far from the rationalistic strategy implying that the naturalistic strategy is somehow flawed, I believe that the former helps support the latter.
iii) A Token-Naturalistic Defence of Self-Reactive Attitudes

A more serious threat to Strawson’s argument is Russell’s account of token-pessimism. Russell’s token-pessimism is compatible with type-naturalism. Token-naturalism states that we cannot permanently give up all instances of certain emotions, such as the reactive attitudes, even if determinism gives us a reason to do so. Russell disagrees with the token-naturalist and believes that we can give up all instances of the reactive attitudes. In response, I will show that token-naturalism is true with respect to the reactive attitudes.

First, we must review the levels of token-pessimism. At the level of events a token-pessimist says that a person is free and responsible except in respect to all her actions. Such a proposed agent seems very strange. She has free will, but this will is not a causal factor in any of her actions. It would be trapped inside her. She would be a puppet on strings, unable to affect what her body does in any way. This disconnection between one’s will and one’s actions is highly implausible, but at least does not require a person to take the objective attitude towards herself. The token-pessimist would continually excuse all such people of responsibility for harmful actions, just as we occasionally excuse people because of accident or ignorance. Yet Russell does not argue that a belief in determinism will make us see every harmful action as an accident, or the result of coercion by another agent, or the result of ignorance, etc. To paraphrase Strawson, this would not result from a belief in determinism, but from universal
goodwill towards others.¹⁵

Rather than argue at this level, Russell places the tokens of token-pessimism at the level of people: "[The token-pessimist] claims only that we can and must cease to entertain reactive attitudes toward any and all individuals who are morally incapacitated [if determinism is true]."¹⁶ According to the token-pessimist, we must retain the potential for having reactive attitudes, but we can and should cease ever to hold these attitudes¹⁷ once we realize that everyone is morally incapacitated. By first looking at the personal reactive attitude of resentment, Strawson indirectly argues against our ability to give up our moral reactive attitudes. I too will indirectly argue that we cannot cease entirely to engage our moral reactive attitudes. I will start by examining the self-reactive attitude of guilt.¹⁸

We sometimes excuse ourselves from feeling guilt for an action despite the harm it did. The first type of excuse, including ignorance, coercion, and accident, allows us to retain our status as free and responsible beings. The second group of excuses consists of global excuses and is more difficult to apply to oneself. We may excuse ourselves globally for certain actions if we acted while suffering a short-term psychological illness, or while under severe stress. Determinism, however, supposedly entails a more permanent excusing condition: if it necessitates that we are morally incapacitated, it proves that we are always so. Thus, according to token-pessimism, although retaining the
potential to feel guilt, a determinist would cease forever to hold that self-reactive emotion.

Since the self-reactive attitudes apply only to whoever has them, a token-pessimist holds that guilt has only one token, the self. The token-pessimist maintains that a person who believes determinism is true and entails her own moral incapacity retains the potential to feel guilt. Nonetheless, such a person can cease altogether to entertain feelings of guilt towards herself, since she believes herself to be unresponsible. Thus, she "can and must" excuse from guilt the only person towards whom she could feel guilty, namely herself.

Wolf has shown that there is no reason for us to take the objective attitude towards ourselves. Yet, whether we have reasons for it or not, can we do so? Wolf's argument does not show it any less rational for us to take the objective attitude towards ourselves, only that it is not more rational. Certainly, one can say, "I am not free or responsible, and therefore, while capable of guilt, I will now stop feeling guilty." Nonetheless, actually ceasing to entertain notions of guilt - that is, ceasing to have a conscience - is more problematic.

First of all, entertaining notions of guilt is different from entertaining a proposition. Since guilt is an emotion, if someone entertains a notion of guilt, then she already feels guilt; this feeling will go away only if she can convince herself that she is not responsible for whatever injury she feels guilt over. So the question is whether anyone can, through a belief in
determinism, consider herself no longer responsible for anything, and, thus, altogether avoid feelings of guilt. The question is whether anyone can cease to feel guilty.

There seem to be two possibilities. A determinist might fail, despite her attempts, to suppress the guilt she feels over her misdeeds. On the other hand, she might successfully prevent herself from ever feeling guilt. In scenario A, upon feeling guilt, she pauses to reflect that she has nothing to feel guilty over, since she is not a responsible being. She attempts to suppress her conscience. One can allay feelings of guilt in specific cases. For example, a person can realize that she is not responsible for a harmful action because of ignorance of its harmful results. Since the action is not a consequence of any malevolent attitudes towards another, she has no reason to feel guilty for it. Yet, she cannot so easily allay her feelings of guilt by telling herself that she is not a responsible agent.

Lacking empirical data about how successfully people allay feelings of guilt, I must pit my intuition against Paul Russell's. An example will help clarify why I think that feelings of guilt are more resilient than Russell supposes. Imagine that John Smith delays buying a new battery for his smoke detector for months, since he always feels too busy. A fire burns his house down and kills his wife and children. John feels guilt over his negligence. His determinist brother, Peter, tries to convince John that nobody is responsible for the deaths of his wife and children. I believe that John will continue to feel guilty even
if he believes in determinism. He may, by holding the objective attitude towards himself for a short time, make himself believe that nobody is responsible for his actions. Nonetheless, this belief is soon overcome by returning feelings of guilt; thus, he still believes himself to be responsible for the deaths of his wife and children.

Guilt may be temporarily suppressed by a brilliant argument showing that nobody is responsible for anything that she does, yet, it returns "as soon as the argument is over." Our conscience is part of our commitment to considering ourselves to be members of the moral community. If our conscience cannot be permanently suppressed, then we have not succeeded in exempting ourselves from the pool of responsible agents. Thus, the above example helps show that token-naturalism is an effective answer to the token-pessimist.

If, nonetheless, a person permanently conquers her guilt, we have scenario B. Here, by eliminating any tendencies within herself to feel guilt, she has eliminated her conscience. People without conscience do exist, such as psychopaths. Certain conditions may even cause people to be without conscience, such as particularly deprived childhoods. Certainly, such folk never entertain notions of guilt, and, thus, have no conscience; they can, at best, debate the rationality of having guilt. These people lack conscience, however, for reasons independent of any beliefs in determinism. Furthermore, they not only lack instances of guilt, but also, unlike normal human beings, lack tendencies
to have guilt; they lack guilt as a type-emotion. Thus we have no reason to believe that those of us who have guilt as a type-emotion can eliminate all tokens of guilt through a belief in determinism.

I think that the only way to eliminate one's conscience, or feelings of guilt, is to silence one's conscience permanently. Whoever does this, however, succeeds in removing her disposition to feel guilt. If she succeeds in this, she does what the type-naturalists deem impossible. Strawson's type-naturalistic argument shows that no theoretical considerations can cause people to remove completely their disposition to feel guilt; thus, scenario B is impossible. This, along with consideration of scenario A, shows that token-naturalism holds with respect to guilt and other self-reactive attitudes.

iv) A Token-Naturalistic Defence of the Other Reactive Attitudes

Next we must determine whether our inability to eliminate our self-reactive attitudes extends to other reactive attitudes, such as resentment and moral indignation, which have more than one object. Some people, such as small children and the insane, do not count as responsible agents. We can eliminate these people from consideration as potential objects of our reactive attitudes. We are able to retain our capacity for resentment, moral indignation, and other reactive attitudes in general without feeling any resentment, moral indignation, etc., towards these people in particular, even when they commit harmful actions. For
such people we do not often experience troubles with reactive feelings being suppressed and later returning to us. We can and do permanently suppress all reactive attitudes towards them. Can we do the same for all people, eliminating all instances of our personal and vicarious reactive attitudes upon believing determinism to be true?

We should remember that, although we permanently suppress our reactive attitudes towards some people, we retain the capacity for reactive attitudes towards normal people. More accurately, we hold these attitudes in reserve for people whom we consider morally capable. But it seems that determinism, if it has any effect upon our behaviour at all, eliminates all people from consideration as moral agents. Nobody for whom to retain these reactive attitudes remains. Thus, we would need a radical change from our current pattern of excusing some people and not others.

Human beings wish to be rational. Russell’s argument likely springs from the belief that people can control their emotions rationally when having them would be inappropriate. Yet, part of the drive to be rational is a desire for consistency. Humans must express guilt when they have knowingly committed a wrong action. They see themselves as responsible even if this conflicts with their belief in determinism. This being so, it is difficult for them to see other people as not responsible beings, unless there is a relevant difference between others and themselves. Determinism fails to provide us with such a relevant difference.

In "Freedom and Resentment" Strawson points out the close
connection among the three types of reactive attitudes. He shows how difficult it is to imagine anyone essentially human who has one or two types of reactive attitudes without the rest. To paraphrase Strawson, a change in our social world that leaves us exposed to the self-reactive attitudes, but not at all to their personal or vicarious analogues, is harder to envisage than the decay of all the reactive attitudes together. Russell goes to extremes to give an example of a situation in which everyone or almost everyone is morally incapacitated: he postulates a rampant brain-disease. A group of such beings is conceivable; one could put some morally incapacitated people together and observe their interactions. Such a group, however, is not a recognizably human society.

We must stretch our imaginations further to picture a society whose members retain only their self-reactive attitudes. This society of moral solipsists, each of whom sees herself as the only responsible being, is even less conceivable than the other, and is also unrecognizable as a human society. Therefore, if one retains her self-reactive attitudes, it seems practically inconceivable that she will not retain her personal and vicarious reactive attitudes. Thus, since a normal person cannot stop feeling guilty for her own actions, then, so long as she believes that another person is relevantly similar to herself she will react with resentment and moral indignation to those of the other.

One can believe oneself the only free and responsible being. There are some people who hold that all others except themselves
are robots, just as some hold that everyone but themselves is unreal. Those who hold these views, however, do not hold them because of a belief in determinism.

As Russell's example shows, one could hold the personal and vicarious reactive attitudes in abeyance, if everyone else succumbed to a morally incapacitating mental illness. This illness constitutes a relevant difference between oneself and others that determinism does not imply. We can see how mental illnesses affect a person's ability to make a realistic assessment of her situation, and to make effective judgements and choices. We can see why she is missing the ability to be part of a moral community, or, as Strawson would say, to participate in ordinary interpersonal relationships. We also find differences between those whom we excuse from moral agency and normal people. This holds even in cases of irrational resentment against objects and natural events, such as the weather. People in modern society tend to personify such natural events, but do not seriously consider a cloud to be a moral agent. Upon reflection, we suppress resentment against the weather, transforming it into mere frustration. We do not blame a cloud for raining as we would a person man for spitting at us. Determinism is different. If it is true, nothing escapes its effects. If a belief in determinism can cause someone to give up her reactive attitudes towards everyone, including those whom she considers relevantly similar to herself, then she should be able to rely on the same belief to give up her self-reactive attitudes. Since she cannot do the
latter, it follows that nobody can give up all reactive attitudes.

If we wish to be consistent we must hold that there are many people, both known to us and unknown, who like ourselves are responsible agents and worthy recipients of our reactive attitudes. This desire to be consistent seems to neutralize our desire to be rational, if the latter requires us to give up all expressions of our reactive attitudes. Thus, our instances of resentment, moral indignation, and other reactive attitudes are as much a necessary part of our lives as are those of guilt.

Yet what about Russell’s analogous token-pessimistic approach to fear? He maintains that, if placed permanently in circumstances in which fear is inappropriate, then, while retaining our capacity for fear, we will feel no fear of anything in particular.

I suppose that a person can believe she is invulnerable, and thus not fear for her life. Perhaps her belief could be strengthened if she survived situations commonly thought to be lethal, like an explosion from an atomic bomb or a hail of bullets. Yet I think that she would express her fear in other ways. She would no longer fear for her life, but might dread other things, such as suffering public embarrassment, or losing a contest, or being alone. And if even the grounds for these concerns are eliminated and she is left with absolutely nothing to fear, then instances of irrational fear might still surface, perhaps in the form of phobias.

Whether or not a person can become totally fearless, I think
that fear is sufficiently unlike the reactive attitudes that the
analogy does not hold with respect to token-naturalism. First,
one can express fear towards many things other than people, since
fear is not tied to interpersonal relations. More importantly,
our reactive attitudes, unlike fear, are associated with seeing
ourselves and others in a certain way. It is not merely because
reactive attitudes are emotions that rational considerations
cannot eliminate them. It is because they are an essential part
of our interpersonal relations, which we cannot give up so long
as we live in a society. People cannot begin to deem themselves
permanently nonresponsible beings, because they cannot stop
expressing their self-reactive attitudes. Similarly, people who
have self-reactive attitudes will be unable to stop having
reactive attitudes towards those whom they see to be relevantly
similar to themselves. Thus, as long as people interact with each
other, it seems practically inconceivable that they will cease
having reactive attitudes.

Of course, many global excuses we make for other people are
matters of degree, thus allowing the others to feel some measure
of guilt for their actions and us to feel some resentment, though
in a reduced form. We retain the reactive attitudes, but
minimally, because of our beliefs about other people. An example
is our limited reactive attitudes towards an eleven-year-old. To
some extent she can be treated as an adult, but in some situations
she must still be seen as an object to be manipulated. Of course,
determinism works differently. If its truth requires us to stop
engaging our reactive attitudes, then this restraint should be absolute. Nevertheless, some pessimists propose a weaker form of pessimism, in which our reactive attitudes are increasingly reduced as we come closer to realizing a rational ideal. Thus, while John Smith will not overcome his guilt over the deaths of his relatives, he may overcome guilt for minor harmful actions.

I have avoided using examples of minor guilt because of humankind’s ability to rationalize excuses for specific actions. People can also ignore their emotions, keeping them on an unconscious level. Thus a person may fail to realize how much she resents a roommate’s sloppy habits until she notices the tooth-grinding and headaches which accompany cleaning up after that roommate. Alternatively, the resentment may build up until she loses her temper.

In any case, I do not believe that one can reduce the reactive attitudes beyond a certain point. One can curb one’s tendencies to overreact to someone’s harmful actions, but it seems that no normal person can reduce the strength of her reactive attitudes beyond a certain level. One might learn not to go into a rage at having one’s property stolen, but would still feel a considerable degree of resentment towards the thief.

v) A Defence of our Rationality

This raises a concern of Russell’s that, if we cannot give up all instances of our reactive attitudes, then we are limited in "our ability or capacity to curb or control our emotional life
according to the dictates of reason." This is rather disturbing. After all, if we were so constituted that we saw ghosts, despite arguments to show it irrational to believe in their existence, it would be distressing that we were constituted to be so irrational. Or consider the principle of induction, which holds that future occurrences of an event type, other things being equal, will be fairly similar to past occurrences. If someone proved that our reliance on induction was irrational, then we would be both wasting a lot of time and money in scientific studies and falsely relying on induction in everyday life.

Let us look at these two examples more closely. Seeing ghosts does not seem to matter so much to our concepts of ourselves as rational beings if we know that they do not exist. We will simply ignore them, walk through them, and otherwise treat them as an optical illusion. If we were unable to do this, then that would indeed be disturbing, but human beings seem capable of making such adjustments. Similarly, if the belief in the existence of ghosts were raised to the level of religious faith, impervious to rational arguments, then the inability to doubt the belief would be disturbing. It is, in fact, disturbing that so many people of so many differing religions are convinced that theirs is the lone true one, and are immune to rational arguments to the contrary. Fortunately, some people are persuaded by these rational arguments.

Nobody, however, could be persuaded by rational arguments to
give up their reactive attitudes entirely. Yet given how much a part of our lives these reactive attitudes are, this is not as disturbing as would be our persistence in seeing ghosts together with the inability to compensate for them. Our reactive attitudes are a basic part of any human society. They colour our interactions with each other in a multitude of subtle and not-so-subtle ways. Therefore, it is not unduly disturbing that we cannot so greatly alter the fabric of our lives, even if some believe it rational to do so.

The example of induction is closer to how our reactive attitudes actually sit with us. We cannot give up using induction to predict future events. It is very likely hard-wired within us as an evolutionary survival trait to be able to predict what is hazardous to our survival and what is not, based upon past experience. In modern science, we sometimes come across cases like quantum physics, in which exact predictions are impossible. Moreover, some of our predictions in everyday life and in science turn out to be false, requiring us to make new ones, hopefully based upon better inductive reasoning. If it were somehow shown to us that induction itself is irrational, however, we would continue to make predictions. Is this disturbing? Perhaps it is to those who wish humans to be completely rational beings. Still, I believe that people are rational enough for almost all practical purposes. It is irrational to expect a creature with a finite brain to be totally rational, able to take all possible factors into account, including its own decision mechanism. This is
especially true with respect to such general and all-pervasive elements of life as reactive attitudes and induction. Therefore, I do not think that we should be too disturbed by our inability to give up our reactive attitudes.

Russell makes another comparison in order to point out the disturbing nature of Strawson's naturalistic argument. He likens us to a group of nonresponsible agents falsely entertaining the idea that they are responsible. If we are in the same situation as these nonresponsible beings, is that not disturbing? I think that it is, and just as we can never be entirely certain that we are not dreaming, so we can never be entirely certain that we are not suffering under a massive delusion. Upon examination, however, this scenario seems unlikely. To test Russell's example, we need to find a group of nonresponsible agents capable of deliberation. Young children are not capable, nor are those suffering from severe stress. As for full-blown psychopaths, they may be able to deliberate over some Schlickian sense of responsibility. Even so, they cannot take seriously the sense of responsibility tied to the reactive attitudes: they lack the very attitudes that make up a normal person's concept of responsibility. Perhaps people suffering under a psychotic delusion consider themselves responsible for actions in their dream-worlds. Are we also all suffering under a psychotic delusion of being free and responsible beings? I cannot prove that we are not, any more than I can prove that you are not dreaming now. Yet if we are deluded, we are at least sharing the
delusion, and not in separate dream-worlds with separate rules. Given that our moral societies function well enough with the idea that we are responsible beings, it is safe to assume that the delusion, if it exists, is harmless. In addition, we are disturbed by the psychotic, not because she considers herself a responsible being, but because the things for which she considers herself responsible have little or no relation to anyone else's reality. She fails to interact with people in the sense we want for us to consider people responsible. Normal people have personal interactions with each other, and we assume that, to a considerable degree, we share a common view of reality. Finally, we hope the psychotic can be cured and become a member of our human society. Yet if we all undertook some extreme treatment to remove our capacity for reactive attitudes, there would be no human society left to join. Thus, if the belief that we are all responsible beings is a delusion, it is one necessary to our functioning as human beings, and, as Wolf has argued, not one that it is rational to abandon.

I think that it is important for people to be able to control their emotions in specific instances. But I do not think it a disaster for human rationality if we are unable to cease entirely to express our reactive attitudes, even if some permanent universal global excuse gives us reason to. Humans are not entirely rational. We are subject to emotions. While we can control them, we will never be rid of them entirely.

In any case, since Wolf has shown that we have no reason to
consider everyone to be without freedom and responsibility, our inability to give up holding our reactive attitudes is not a threat to our dignity as rational people. Thus, those who find it disturbing that we are not able to give up our reactive attitudes, even if it is rational to do so, can take some comfort in our having no reason to give them up. This is good for us, since Wolf also shows that we have good practical reasons to retain them. If we took the objective attitude towards everyone, we would be in a cold world of human isolation in which such important facets of society as friendship and love are drastically reduced in meaning. We would all become human calculators that lacked many human feelings and that were unable to relate meaningfully to anyone. Worse still, if we gave up taking attitudes entirely, we would lose our sense of ourselves as persons, and all of our values. 

"A world without reactive attitudes would be a tragic world of human isolation; a world without reactive attitudes or the objective attitude would be a bleak, blank world of human brutes." 

Strawson's argument has some imperfections, however. Russell correctly states that the rationalistic strategy is flawed with respect to global excuses, since Strawson confuses moral incapacity with mere abnormality. From examining the cases of moral incapacity, however, I think it has been shown above that no effective global excuse can be universal and permanent, since it cannot permanently include the self. Therefore, while we may be determined, and while determinism may entail that we are all
"unfree and unresponsible", we cannot give this as an excuse for people's actions. Neither is it rational for us to give such an excuse, either pragmatically or in the sense of what it is rational to believe in the privacy of one's own study, as Downie wants to suggest.\(^3^0\) Therefore, the flaw in Strawson's rationalistic argument is not a serious one, and the objections raised so far to his general argument have been answered.
1. I gave a talk formed from part of this chapter at a Philosophy Department Graduate Student Colloquium at the University of British Columbia on Dec. 5, 1994, and am grateful to those who attended, particularly Paul Russell, Adam Constabaris, Richard Johns, Michael Fleming and Alan Richardson, for their helpful comments.

2. It sometimes seems that Strawson is sympathetic to determinism, especially when he mentions the incoherence of contra-causal freedom. However, he takes pains to remain neutral when discussing determinism, claiming not to know what it is (Strawson, "Freedom and Resentment", 71, 95).

3. Wolf, Susan, "The Importance of Free Will" in Mind 1981 vol.XC, 386-405. This question is also addressed in a similar fashion in J.E. Llewelyn's "The Conceivability of Pessimistic Determinism" in Analysis vol.27, no.2 (new series no. 116), Dec., 1966, 39-44, but I believe that Wolf's answer to the problem is more complete.


8. Wolf, "The Importance of Free Will", 400.


17. Russell sometimes argues that the token-pessimist’s viewpoint is that people are able to cease entirely to experience an emotion, such as fear or a reactive attitude, while still being "prone or liable to [it]" (294), and sometimes seems to argue that the token-pessimist declares that we cannot "reasonably or appropriately entertain or engage these attitudes" (296). The latter interpretation seems to be an attack upon the rationality of holding the reactive attitudes. It is, I think, answered by Wolf. The former interpretation I hope to answer in the following pages.


19. We can make an exception for psychopaths, who are incapable of guilt but are rational enough to ask the question "Why should I feel guilty about my actions?" Since psychopaths are incapable of feeling guilt, however, or of understanding what guilt feels like, I think that their attempts to entertain the notion of guilt are significantly different from those of most people. For psychopaths, to entertain a notion of guilt is the same as to entertain the truth of a proposition. If Russell means that normal people "can and must" cease to entertain a notion of guilt in this sense, then he is arguing for the irrationality of holding the reactive attitudes, not for our ability to cease to hold these attitudes.


21. Galen Strawson describes the possible use of Buddhist meditations to experience the lack of freedom that incompatibilist determinism entails, in "On 'Freedom and Resentment'", 98-100. Wolf shows that there is no rational reason to do this, and I think that most people’s desires to retain their sense of identity provides a powerful reason to refrain from these meditations. Moreover, although I do not wish to begin a detailed criticism of the Buddhist religion in this thesis, I think that such a practice cannot be entirely successful. I agree with Galen Strawson that the practice, to the extent that it succeeds, will make people inhuman, though not necessarily inhumane. The meditations seem by their nature to be solitary and thus require that one severely limit personal interactions with others. Furthermore, insofar as the achievement of a partial objective attitude is possible, it takes years of practice; one’s self-reactive attitudes cannot be reduced by an "effortless, rationally motivated, self-directed intellectual fiat" (98). Thus, a belief in incompatibilist determinism alone will not suffice.
22. For a detailed examination of psychopaths, see Robert D. Hare’s *Without Conscience: The Disturbing World of the Psychopaths Among Us* (New York: Pocket Books, 1993).


27. Paul Russell was responsible for the suggestion of seeing ghosts and of the deliberating nonresponsible agents. Induction is referred to in Strawson, "Freedom and Resentment", 94, note 1, but was also mentioned in conversation with Richard Johns.


29. Wolf, 400-1.

30. Downie, 36.
i) Predictability and Determinism

I will consider one of Strawson's critics in more depth. A.J. Ayer holds a view similar to Downie's. Both declare that the objective attitude has more in common with determinism than Strawson admits, and that there are reasons to believe that reactive attitudes are inappropriate. Yet, Ayer's criticism differs from Downie's because Ayer believes it possible to give up the reactive attitudes. By examining Ayer's essay "Free Will and Rationality" I will bring Strawson's views of reactive attitudes into sharper focus.

Strawson's states that we can sustain the objective attitude towards another only when we have special reason to do so. This reason derives from viewing an agent as being in some way abnormal, and so not a member of the moral community. In light of Russell's comments in the last chapter, it is better to say that we see the agent's moral capacity as relevantly different from our own. Thus, she is an inappropriate object for our reactive attitudes. Since in all known societies this relevant difference concerns abnormality, I will not dwell on this point. In any case, Ayer acknowledges the force of our global excuses in current practice. He holds, however, that our ability to take the objective attitude for these reasons does not show that we cannot take a sustained objective attitude for other reasons, despite Strawson's assertion that people cannot give up their reactive attitudes.
Ayer allows that as societies now stand, it is probably impossible for us to give up our reactive attitudes merely because of belief in a general thesis concerning human behaviour, such as determinism. He admits that the determinists of his acquaintance experience reactive attitudes as strongly as other people. Still, no scientists have studied what would happen if determinism were well-known. Alternative conjectures to Strawson’s account may be practically conceivable. If any are, then Strawson’s assertion may be false.

Ayer attempts to lay the groundwork for such an alternative, by showing that the content of determinism is important to the theory’s influence on human behaviour. Ayer selects two points about determinism that he considers to be significant. First, if determinism is not vacuous, it must state that human action is subject to manageable laws. By manageable, Ayer means that we should be able to use generalizations of human behaviour to explain and predict accurately, or, with strong probability, "every facet of human behaviour". This makes determinism a strong thesis, and many professed determinists are not determinists in this sense. Nevertheless, Ayer finds no a priori reason to reject this version of determinism.

Ayer’s second point is that determinism requires a person to be unable to act otherwise than she does. He maintains that this inability is proportional to how well we can predict or explain actions accurately, or with a high degree of probability, given "a set of particular facts and a set of well-established
hypotheses which together entail or at least make it highly probable that the person act as she did. This relates ability to do otherwise to the state of knowledge of the observer. Ayer postulates that we can consider a person able to have acted otherwise only as long as nobody can provide the requisite set of facts and hypotheses.

With a strategy similar to Downie's, Ayer attempts to connect determinism to the objective attitude using predictability. Employing these relationships, he states that taking the objective attitude is not primarily dependent upon Strawson's global excuses, but upon the ability to predict someone's actions exactly and, thus, show that she could not have done otherwise. Ayer's determinism, if true, requires that such predictions can be made.

Despite his aims, Ayer fails to tie his high standards of predictability to the objective attitude. First, nobody has the scientific knowledge that Ayer's determinism requires in order to predict accurately what people will do. When we take the objective attitude towards people, we are not situated to make exact scientific predictions based on an array of facts and well-established hypotheses. Rather, we make general, fallible predictions of behaviour. Ayer describes our ability to predict the actions of members of our moral community as follows: "They are statements of tendency which are not thought to be sufficiently strong or far-reaching to make the actions which they govern unavoidable." Yet, the same description applies to attempts to predict the actions of nonagents.
In reply to Ayer, Strawson mentions that we do not use scientific laws when we take the objective attitude. He adds that the vague laws of human nature we have apply to all people equally, regardless of our attitude towards them. When we take the objective attitude, we usually focus on determining what causes are likely to produce certain effects; that is our usual reaction to complex objects. We do not gain with this focus, however, the ability to predict behaviour with increased accuracy. In fact, we can have more trouble predicting a lunatic's actions than those of a sane person.

Nor do we employ a prediction process qualitatively different with nonagents than with normal people, contra Ayer. When a psychiatrist treats a patient, she has no table of scientific laws from which to judge which stimuli will elicit which responses. Rather, she judges from experience the treatments most likely to be effective. When the psychiatrist visits her parents, she may similarly judge from experience the gifts likely to please them. The difference here is that she cares about more than possible responses to possible stimuli; she cares about how her parents feel about her and, to a lesser extent, how they feel about others. If her parents insulted her, she would be shocked and resentful; if her patient insulted her, she might be shocked, but probably not resentful. In both cases she would try to find out why the insults occurred and how they could be stopped, but in neither case would she use the kind of scientific laws Ayer demands.
Ayer recognizes that determinists do not have the requisite hypotheses to predict people's actions. He believes that this lack of a working theory of human physiology and psychology alone allows them to have reactive attitudes and interpersonal relationships. By Ayer's own observer-oriented definition of "could have done otherwise" contemporary determinists can thus say that normal people could have acted in another way. Yet, Ayer must hold that people to whom we take the objective attitude are so simple that we can use current scientific knowledge to accurately predict their actions. We have no evidence for this, or for thinking that an immature or damaged mind is so easy to analyze and predict. Even if it is so, Ayer fails to explain why, if it requires such extensive knowledge, we are briefly able to take the objective attitude towards normal people.

ii) Science and the Objective Attitude

Ayer has not sufficiently argued that to adopt an objective attitude requires us to know the causal nature of others in great detail. Yet he thinks we might discover a theory one day that enables us to analyze the causal patterns of everyone. If so, we might have to give up our reactive attitudes. Therefore, I will consider Ayer's account of how his version of determinism can come into general knowledge.

Ayer begins by pondering what would happen if "the requisite physiological or psycho-physical theories were developed and we could use them in everyday life to make mainly accurate
predictions". Given its record people tend to be optimistic about what science will one day be capable of. After all, science has allowed discoveries that seemed impossible a century earlier. Can we be sure science will never develop the theories Ayer describes?

Strawson thinks we can. Reactive attitudes are adopted in a variety of situations, some of which are very complex. Moreover, the attitudes themselves are sometimes quite subtle. It would be "a practical absurdity" to imagine anyone establishing the useable, accurate laws that Ayer wants. These laws must correlate all of the subtleties of the interrelations between our thoughts, feelings, and actions with a causal system of physical brain-states, or whatever other exact scientific terms one wishes to use. Strawson holds that this is too much to expect, even of science. This is an empirical statement; it may be proved false. Still, I think it is a reasonable assumption. As Daniel Dennett notes, many chaotic systems in nature, such as the path of a planet in a chaotic orbit, or even the path of a pinball after hitting more than twenty posts in a few seconds, are absolutely unpredictable to us. The workings of the human mind, being even more complex, must also remain unpredictable. Also, a person's knowledge of the laws purporting to describe her own behaviour might modify that behaviour, making prediction prohibitively difficult. Therefore, I think Strawson correctly supposes Ayer's example to be practically inconceivable.

Ayer admits that it may be too fanciful. His second example
is more believable. Ayer proposes that scientists might develop a theory of conditioning "to the point where it became possible to implant desires and beliefs and traits of character in human beings, to an extent that it could be deduced, at least in fairly general terms, how any person who had been treated in this way would most probably behave in a given situation". Ayer believes that, if this conditioning was practised on us, and if everyone was given sufficient knowledge of the conditioning process to predict or explain "almost everything that anybody said or did", then we would forego our reactive attitudes.

Currently we can condition people to a degree. Brainwashing and advertising are both testaments to how well a conditioner can influence people to think and act as she wishes. Given how complex the human mind is, though, it is very unlikely that such techniques can control everything that a person says or does. People face a much lower level of conditioning from advertisers and brainwashers, although the latter have considerable power to influence one's thoughts. Ayer's conditioning is absolute, controlling every single aspect of a person's life. If anyone develops this form of conditioning, I think that the human mind's complexities would be lost.

Strawson shares this view, holding that, if Ayer's conditioning were practiced, "the patterns of human behaviour would be simplified out of all recognition". Dennett describes the required process of conditioning as crude and presumably "powerful enough to override [a person's] delicate, information-
gathering, information-assessing, normal ways of coming to have beliefs and desires". Ayer's conditioning process would, as Ayer predicts, make us give up our reactive attitudes. But no recognizably human society will remain in which to adopt reactive attitudes or have the personal interactions that are tied to them. Although our reactive attitudes are an essential part of us, they must be nurtured in a human society. If that human society is abolished by Ayer's conditioning, then the reactive attitudes will likely be abolished with them. Thus, Ayer's example reveals the importance of human society to the reactive attitudes, but it fails to show that the former can exist without the latter.

iii) The Applicability of Desert

In the last chapter we saw that Wolf has shown that even if we could dispense with our reactive attitudes, there are powerful reasons to retain them. Nevertheless, Ayer proposes a sense of rationality according to which determinists are irrational to hold the beliefs in human responsibility to which the reactive attitudes lead.

Strawson maintains that no general theory can make us lose our reactive attitudes. Ayer interprets this to mean that "no thesis which applies to human conduct indiscriminately can pose such a threat". Ayer also interprets a person who holds the objective attitude as one who responds to or judges human behaviour merely "from the point of view of its causes and effects". Thus, Ayer holds that Strawson thinks it practically
inconceivable for us always to judge human behaviour merely in causal terms. This is close to saying that it is practically inconceivable for us to respond to each other as if determinism were true, even if it is.

Yet to respond so may be rational. Ayer declares that in our reactive attitudes "there is a tacit or explicit reference to the concept of desert".\textsuperscript{25} Ayer ties this concept to a belief that the object of one's reactive attitudes could have done otherwise.\textsuperscript{26} Ayer believes that desert and, hence, the basis for the reactive attitudes, is threatened by determinism. Determinism holds that, given all causal antecedents exactly as they are, nobody can act otherwise than she does.\textsuperscript{27} Yet Ayer believes that a contra-causal ability to do otherwise is required by the concept of desert. This concept thus has no application if determinism is true. Since Ayer believes responsibility to be founded on desert, responsibility will also be an empty concept. And, if responsibility has no application, then we are irrational to hold the reactive attitudes, which commit us to a belief in the general responsibility of people for their actions.\textsuperscript{28}

This argument resembles Downie's, which was raised in chapter III and answered in chapter IV. There Wolf showed that it is no more irrational for us to keep our reactive attitudes than to give them up. Of course, a being without reactive attitudes may have no reason to adopt them, except the instrumental concerns which Strawson describes. Yet, since we already have and cannot excise our reactive attitudes, we are not in such a being's place. Thus,
we are not irrational to consider people generally responsible, whether they have contra-causal freedom or not. If Ayer's definition of responsibility is empty, then we should find one that is more serviceable.

Ayer raises an important concern, however. He tries to produce a definition of responsibility which is more than a correlation with our reactive attitudes. His definition is flawed since it relies on a concept of desert that is empty if determinism holds. Nonetheless, we probably would be better off if we had a definition of responsibility. People can be uninformed of the facts and, thus, hold reactive attitudes in situations for which they are inappropriate. We could use definition to help clarify when the reactive attitudes are and are not suitable.

On the other hand, perhaps we do not need a definition of responsibility. A sufficiently well-informed person will have reactive attitudes in all the right situations and no others. Thus, we should be able to correlate responsibility with the grounds of these reactive attitudes.

Unfortunately, it we can not safely rely on always having the relevant facts. Furthermore, we sometimes encounter cases in which people disagree on whether the reactive attitudes or the objective attitude is appropriate. In the next chapter Gary Watson gives an example of one particularly controversial case. If we had a definition of responsibility, we could use it to help resolve such cases.


5. Ayer, 6.


10. Ayer, 7.


22. Dennett, Elbow Room, 34.


26. Ayer allows (Ayer, 6) that this sometimes indicates that we believe a person able to have altered his behaviour at an earlier time, even if he could not have acted otherwise at the time of his blameable action. This lets us blame drunken drivers who kill people in an accident.

27. Ayer, 3.

Chapter VI: Watson's Partial Definition of Responsibility

i) Global Excuses Taken Individually

In this chapter I reveal the clarifying power of a definition of responsibility, but I will not attempt to define it yet. First, I must consider whether defining responsibility undermines Strawson's argument. Strawson does not believe it important to find a strict definition of the reactive attitudes. In "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme," Gary Watson argues that, according to the Strawsonian model, responsibility cannot be defined independently from our reactive attitudes. Strawson holds that our reactive attitudes are immune to total elimination by mere theoretical considerations. Watson concludes from this that reactive attitudes alone form the concept of responsibility. Theoretical definitions have no place. Thus, responsibility is as much a brute fact as the reactive attitudes themselves; any attempt to define responsibility independently of reactive attitudes will, according to Watson, miss this vital point.

"It is not that we hold people responsible because they are responsible; rather the idea (our idea) that we are responsible is to be understood by the practice, which itself is not a matter of holding some propositions to be true, but of expressing our concerns and demands about our treatment of one another" (emphasis his).

Strawson holds that the reactive attitudes are basic to our nature. "[They] are precisely the correlates of the moral demand in the case where the moral demand is felt to be disregarded. The making of the demand is the proneness to such attitudes"
(emphasis his). This does not mean, however, that he must hold that a separate propositional definition or analysis of responsibility is impossible. Just as a baseball player can catch a ball without doing a complex physics equation to ascertain the ball's location and velocity, so we can use our reactive attitudes to hold people responsible without having a propositional definition of responsibility. Nevertheless, we can search for theoretical equations or definitions in both cases.

To say that we might find a definition of responsibility for Strawson's framework of reactive attitudes is one thing. Showing the need for such a definition is quite another. I will not define responsibility in this chapter. I believe, however, that Watson shows how a definition can help clarify cases in which the suitability of the reactive attitudes is in question.

Watson notes that children, psychotics, and others who are excused from being objects of our (full) reactive attitudes can still show hostile attitudes towards us. Our reactive attitudes are thus "sensitive not only to the quality of others' wills, but depend as well upon a background of beliefs about the objects of those attitudes". Watson clarifies some of these beliefs by investigating why our reactive attitudes are reduced or negated by global excuses. According to Watson, "many of the exempting conditions involve explanations of why the individuals display qualities to which the reactive attitudes are otherwise sensitive" (emphasis his) and, therefore, why in these cases reactive attitudes are inappropriate, even if an injury has
occurred.

By examining Watson's findings I will show what some of the conditions of responsibility are. Watson looks first at children. He believes that adults have moral understanding, which children have to learn. They discover how others are hurt, both physically and psychologically, and "the various subtle ways in which that may be done". They also learn what is expected of them as adults, namely, a reasonable regard for others. Since this learning process is gradual, our reactive attitudes to a child will be less inhibited as the child grows older and, presumably, has greater moral understanding. Of course, other things besides moral understanding are necessary parts of responsibility. Nevertheless, moral understanding seems to be one of its preconditions.

There are others. Some global excuses apply to people under temporary but extremely severe stress; such people are said to be "not themselves". We require people to "be themselves" when acting if they are to be appropriate recipients of the reactive attitudes. Watson suggests that the reactive attitudes are related to moral demands that we make of one another. If we address a moral demand to a person who has acted while under stress, however, we address the true moral self, who has the proper attitude towards people even though she was not in control. It is as if the moral self was overwhelmed by a second self who controlled the person's actions while the person was under stress. When the period of stress is over, the second self flees, leaving
us with the "innocent" primary self. Furthermore, the second self
would seem more worthy of the objective than of the reactive
attitudes, since a person while under stress acts in a manner that
is not wholly sane. "Insofar as resentment is a form of reproach
addressed to an agent, such an attitude loses much of its point
here...because [the true self] repudiates such conduct as well". Watson follows Jonathan Bennett's interpretation of reactive
attitudes as a form of moral address. I will examine this in the
next chapter.

Strawson says that the infrequency of the stressful periods
allows the reduction of our reactive attitudes. "We normally deal
with [the agent] under normal stresses; so we shall not feel
towards him, when he acts as he does under abnormal stresses, as
we should have felt towards him had he acted as he did under
normal stresses." If a person were frequently to act as if under
stress, we would likely increase our reactive attitudes towards
her, partly because we would blame her moral self for not being
able to deal with the stresses. Sometimes people are overwhelmed
by the stressful situations they encounter. We expect them to
deal with these situations eventually, however, and to participate
once more in the moral community as full agents. On the other
hand, if someone suffers a series of severe stresses, she might
go insane. She then fall into a more permanent exempting
category.

Watson raises some problems related to thinking of someone
as not being herself. One problem is how to articulate what a
moral self is. This will be easier once a definition of responsibility is found. Another problem is that we usually consider the person to be one being, instead of dividing her into a moral self and a stressed self. Thus, we may prefer to see responses under stress as "reflections of our moral selves - namely reflections of the moral self under stress". However we look at stress, we require the agent to have a certain degree of control over her actions, choices, and character for reactive attitudes towards her to be appropriate. This is true whether she loses control to "another self" or simply loses control. Whatever self-control is, it is exhibited by and expected of most adults.

Perhaps we inhibit the reactive attitudes in cases of stress because we recognize that people are imperfect; they can be overwhelmed by great shocks, such as a divorce or the death of a loved one, and lose their self-control. They may thus fail to give the consideration for others that we otherwise require. So it appears that we limit reactive attitudes in these cases because we know that severe traumas can overwhelm people’s self-control. There must still be societal consensus of what counts as a sufficiently traumatic shock, and this may vary among societies. Furthermore, defining self-control exactly may be difficult. Nevertheless, self-control of some sort is another general component of an account of responsibility.

Some cases not analyzed by Watson are worth examining. The psychopath, in the extreme a being incapable of feeling guilt, is a complex case. At first, the reason we abandon reactive
attitudes towards her seems the same as for doing so towards children. Both psychopaths and children lack moral understanding. An added complication in the case of psychopaths, however, is that they are as intelligent as normal people. They can understand society’s laws, can learn which actions are acceptable and seem to be in control of themselves. They do not understand why harmful actions are wrong, but they can at least realize that such actions are called wrong. While they have no moral sense, they are often able to act as if they do.

When a psychopath commits a particularly horrible crime it gets even more complicated. Our reactive attitudes come into play, and although some people may disengage their reactive attitudes when they learn that the criminal is a psychopath, others may continue to blame her. Watson notes that a person plainly less easily disengages reactive attitudes when "the knife is at one’s own throat", but we can also feel moral indignation when a psychopath harms someone whom we do not know. These reactive attitudes may disappear when we realize that the wrongdoer is a psychopath, or they may not. Is this a sign that people as a whole have more social evolving to do? Perhaps, but this lack of social evolution has an explanation. Psychopaths are quite adept at mimicking humans who have a moral sense. While tests have been developed that can determine whether someone is a psychopath or not (and some people are apparently able to recognize a certain coldness in a psychopath’s eyes) in ordinary life people are easily fooled by psychopaths. Thus, with respect
to psychopaths, our reactive attitudes are not always accurate. Psychopaths can trick us into believing that they are part of our moral community; we then react to them as such when they act wrongly. Another reason why we do not as readily adopt the objective attitude is that it often deals with treatment and control, but there is no known method of successfully treating a psychopath. Over the centuries, people have differed in opinion over whether psychopaths are mad, and not part of our moral community, or evil, and a part of it.\textsuperscript{14} Since in the extreme psychopaths are not prone to guilt or moral indignation, Strawson would probably judge them as mad, and would exclude them from consideration as responsible agents.

Another case of exemption is that of psychotics\textsuperscript{15}. Psychotics are trapped in their own fantasies, and cannot interact with the real world. One reason for disengaging our reactive attitudes towards them may be that they lack moral understanding. More importantly, they lack a general understanding of what goes on around them. A person believing she is in ancient Rome may act like a good Roman citizen, and, thus, have a moral sense that is appropriate to her concept of Roman society but inappropriate in a modern society. Interacting with her as a member of our moral community would be difficult or impossible. More suitable, it seems, are treatment and control, the hallmarks of the objective attitude.
ii) The Unfortunate Childhood

A final category to investigate concerns the effects of a peculiarly unfortunate childhood. Robert Harris is referred to by Watson as an example. Harris's case demonstrates why a definition of responsibility may be useful. Harris was thoroughly vicious, cruel and evil. He killed two teenagers and then, laughing, ate their hamburgers. He claimed to have chosen "the road to hell." Thus, he seems a perfect candidate for the reactive attitudes.

On the other hand, Robert Harris was certainly unfortunate in his formative circumstances. He was beaten by both his parents and was never shown any affection. As a youth, he was raped repeatedly in a federal detention centre. This physical and mental abuse seems to explain how Harris became a monster. Although, as Watson points out, we continue to adopt reactive attitudes towards him, we can now see Harris as a victim as well as a villain. Watson believes that trying to see him both ways leads to an incoherent viewpoint. Does Harris count as a moral agent, or should his abusive childhood exempt him from our reactive attitudes?

We may need to modify Strawson's viewpoint to accommodate people like Harris. I think, however, that Harris is a candidate for the reactive attitudes, even if he rejected the values of his moral community. Extreme evil, as Watson says, cannot be its own excuse. Yet, we also see Harris as a victim. This can be explained as an effect of feeling reactive attitudes towards all
those who had abused him throughout his childhood. We see how Harris became the way he was, and can even pity him. Still, we refuse to excuse him of his crimes. Our moral indignation, in this case, overrules our pity.

As Watson notes, Harris's case may exemplify one form of moral luck. The problem of moral luck points out that, among other things, many of our character traits are the result of formative circumstances over which we had no control. Could we have avoided becoming as evil as Harris if we had had a childhood as bad as his? I will address the problem of moral luck in a later chapter. As a brief answer, people have survived worse childhoods than Harris's without becoming moral monsters, which at least shows that a horrible childhood does not automatically make one a nonresponsible agent.

It is significant that Harris is not excused from the reactive attitudes, even though he had "peculiarly unfortunate formative circumstances". Strawson lists this exempting condition among the possible excuses that partially or wholly disengage our reactive attitudes. This condition is troublesome because its meaning is ambivalent. Does "peculiarly unfortunate" mean extremely unfortunate in formative circumstances, making Harris's case one that should disengage our reactive attitudes, or does it mean that the formative circumstances are unfortunate in a peculiar way? Certainly, if the formative circumstances lead the person to lack moral understanding (unlike Harris, who seems to have understood that what he did was evil, claimed to have
chosen to be evil\textsuperscript{21}, and was, in the end, capable of guilt\textsuperscript{22}), then we do not consider her to be a moral agent. But many sorts of unfortunate upbringing do not invite the reduction of our reactive attitudes, although they may invoke our sympathy, and cause us to adopt reactive attitudes towards those responsible for abusing them. The Harris case is an example of our retaining our reactive attitudes despite our knowledge of a horrible childhood. Therefore, contra Strawson, I do not believe that peculiarly unfortunate formative circumstances do or should cause us always to repress our reactive attitudes. When we do abandon them due to someone having lived through such unfortunate circumstances, it is because the circumstances caused the person to fall into another exempting category. They may, for example, cause someone to become psychotic.

The above cases provide rough elements that help to form a definition of responsibility. They may help decide when reactive attitudes are either appropriate or not. This sort of internal critique of the reactive attitudes is compatible with Strawson's thesis:

"It is, of course, frequently in order to declare that some instance of an emotional reaction is inappropriate or even irrational."\textsuperscript{23}

"We can see where the limits of our proneness to [reactive] attitudes tend to fall, we can understand why they tend to fall where they do and we can find room for the idea of criticism, of appropriateness and inappropriateness, in particular cases."\textsuperscript{24}

After examining the Harris case we can modify Strawson's list of exempting conditions. By examining these conditions in the
above cases, we gained a partial account of responsibility. While being a child or a psychotic or a psychopath or temporarily under stress remain as excusing conditions, not all forms of unfortunate formative circumstances constitute legitimate excuses. The unfortunate formative circumstances must lead the person either to have a limited moral understanding, or else to fail to fulfil another condition of responsibility. It is likely that a more complete account of responsibility will be even more useful in deciding when the reactive attitudes are inappropriate. Of course, these attitudes should not be overruled by an abstract definition. If such a definition conflicts with our reactive attitudes, then either an explanation for each type of conflict is necessary or the definition will be judged wrong and a new one will replace it. Assuming that a good definition or analysis of basic conditions is found, it will help to clarify the internal criticisms that we make concerning to whom we should apply our reactive attitudes, without criticizing the reactive attitudes as a whole.

iii) Living Counterexamples: Einstein, Gandhi, and King

Watson has other concerns besides finding an account of responsibility. The words and actions of some people do not seem to fit Strawson's theory. Albert Einstein, being a determinist (and having a sense of humour) declared that he did not hold himself or others to be responsible beings. Furthermore, he asserted that he was able to distance himself from all other
people. He seems to imply that he has adopted the objective attitude towards everyone.

"Everybody acts not only under external compulsion but also in accordance with inner necessity...This realization mercifully mitigates the easily paralysing sense of responsibility and prevents use from taking ourselves and other people all too seriously...[With respect to society, family and friends] I have never lost a sense of distance and a need for solitude". 25

If Einstein is accurate in his assertions, then either he is not a normal human being or Strawson's theory is false. I think, however, that Einstein is not quite accurate in his assertions about himself. I believe that he did not take himself or others "too seriously". Yet, Einstein's feelings merely mitigated, or reduced, his sense of responsibility. A belief in determinism, if it has any effect on our feelings of responsibility, requires us to negate them completely. I believe that the answer lies in Einstein's emphasis on humour being given its due.

I can only speculate on why Einstein refused to take people seriously, but I have shown in earlier chapters that one cannot simply negate one's sense of responsibility by an act of will. Perhaps Einstein is compensating for people's tendencies to overreact to some situations, especially to "offences" that are more breaches of etiquette than intentions to harm others. If so, it explains why his sense of people's responsibility was mitigated, instead of entirely removed. Many people have claimed that, if determinism is true, then people are not really responsible for what they do. Einstein might upon hearing this
claim have thought that this is why he did not take people seriously. Perhaps similar reasons hold for others who claim that, because determinism is true, they will not hold people to be responsible beings. In any case, if some sort of "false consciousness" does not operate in these situations, then Strawson's essay needs to be reexamined. Strawson declares that we cannot give up our sense of responsibility, which is the expression of our reactive attitudes. He is wrong, however, if people can wholly give up such attitudes by an act of will.

Einstein's claim that he never belonged to his friends or society with his whole heart, and that he sought solitude, is not necessarily an example of the objective attitude. It is possible to wish at times to be alone, or to keep largely to oneself, without giving up one's reactive attitudes. Someone left alone on an island from early childhood may not have any reactive attitudes. Yet people like Einstein who are raised in society will still adopt these attitudes.

Watson has other concerns. Should it be an ideal of ours to give up the negative elements of our reactive attitudes, such as resentment and aggression, even if it seems practically impossible? Many people hold universal goodwill towards others to be an ideal for humankind, even if this is seldom if ever practiced. Do we have to give up our interpersonal interactions to realize such an ideal? Watson believes that Strawson says as much in his own view of punishment. Strawson holds that our negative reactions to others are closely tied to our reactive
attitudes.

"Indignation, disapprobation, like resentment, tend to inhibit or at least to limit our goodwill towards the object of these attitudes...the preparedness to acquiesce in that infliction of suffering on the offender which is an essential part of punishment is all of a piece with this whole range of attitudes." 26

Retributive elements seem to be tied too closely to our reactive attitudes for us to hold the latter while maintaining universal goodwill. Yet, some people, such as Martin Luther King and Gandhi, seemed to have no ill feelings towards anyone, even the people whom they held responsible for various wrongs. 27

We might speculate that these people are forgiving of all actions. In Watson’s description of King, Gandhi and Einstein, they show a similar lack of illwill towards others. Yet, King and Gandhi seem to have retained the reactive attitudes of resentment and moral indignation. King and Gandhi did more than state that certain people were responsible for certain crimes; they made moral demands that the wrongs be righted.

As Einstein states his case, there seems to be no demand addressed to people in such a manner. Thus, his case might be considered closer to universal forgiveness. This view is problematic. Einstein claimed that people were not ultimately responsible for their actions. Yet, to forgive is still to hold people responsible. We grant forgiveness because the people we forgive express regret over attitudes which led to the wrong action, and they make a commitment not to have such unacceptable attitudes in the future. 28 This points towards another reason
that none of the three examples fit the model of universal forgiveness. The people Einstein, Gandhi and King are supposedly forgiving are not necessarily sorry for what they have done, nor do they forswear future wrong actions. Thus, forgiveness is inappropriate in these cases. We must find another description for Einstein's statements and King's and Gandhi's actions.

To address the case of Einstein first, we should examine the requirements of a doctrine of universal goodwill. Einstein mentioned a conceptual isolation from the rest of society, and it may be thought that isolation helped him to take people less seriously. It is not obvious, however, that one has to be isolated, either physically or conceptually, in order to feel universal goodwill. In fact, goodwill to others borne in isolation misses the point. Universal goodwill towards others is an ideal that is sought in societies, not in isolation. It is supposed to help people live in harmony, not in seclusion. It is the perceived lack of goodwill towards others which triggers our reactive attitudes. Therefore, isolation does not seem a necessary or desirable part of the ideal of universal goodwill.

In any case, if we are able to give up our negative reactive attitudes through isolation, we will also lose the positive reactive attitudes, such as gratitude and some kinds of friendship and love, through the same isolation. The loss of our interpersonal relationships is a heavy price to pay to realize the ideal of goodwill towards all. I think, however, that we need not isolate ourselves in order to feel a large amount of goodwill
towards others. And, unless we practice such drastic measures as total isolation, I know of no means by which we can give up our reactive attitudes as a whole.

Yet perhaps people can give up their wish to harm others while retaining their reactive attitudes. Are King and Gandhi such people? I think not. While they did not advocate violence, their actions indicate that they did approve of wrongdoers suffering moral censure and public disdain. Therefore, they may have differed with others only in the type of punishment deemed appropriate, not the wish to punish. Society seems to be slowly evolving in the direction of reducing punishment. In western societies extreme physical torture is no longer viewed as acceptable punishment. But a wish to reduce overreaction to harmful attitudes does not entail a wish to cease punishing wrongdoers. We may merely be seeking a proper level of punishment.

Let us assume that there are people even more free of a wish to punish than Gandhi and King. These people will have no wish to harm anyone. Could they still retain reactive attitudes, while having no wish to punish wrongdoers? Strawson says that reactive attitudes such as moral indignation and resentment only tend to inhibit our goodwill towards the person to whom we feel such attitudes. This leaves open the unlikely possibility of feeling resentment and moral indignation towards another without the accompanying wish for the other to be punished. And if the wish to punish is not necessary to the reactive attitudes, then the
"preparedness to acquiesce in that infliction of suffering on the offender" is also unnecessary. It is very easy for people to want others to suffer for what they have done. This wish might be basic. It might also come from a combination of a wish for others to realize that what they have done is wrong, together with a belief that they may learn that their actions are wrong if they both are the objects of moral indignation and are punished. It is, however, at least conceivable that one could hold people responsible for their actions, but merely desire that the wrongs be righted, or at least for the wrongs to stop, rather than want to cause the wrongdoers to suffer as well.

It seems odd to use words like "resentment" and "indignation" and yet hold that they do not entail that others undergo some negative experience. I believe that many people think these emotions contain a wish to harm others because the emotions are frequently accompanied by such a wish. We are not saints, and when someone does wrong we usually feel not only the reactive attitudes, but the hot form of these attitudes. The latter consists of seeing the offender as an enemy to ourselves and the rest of society, to be dealt with mercilessly. If we were all totally benevolent we might have no wish to punish people whom we held responsible for crimes, but simply wish them to right the wrongs they have committed. But, if we were all totally benevolent, then nobody would show improper attitudes towards anyone in the first place. A totally benevolent society would have no need for punishment, which, in actual societies, is the
expression of our reactive attitudes.

While it may be an ideal to be like Gandhi or King, it may not be an ideal to be totally benevolent in the above manner. If we had no desire to make offenders suffer, we would have no desire to make them feel guilt, as this is a form of suffering. We would only retain a desire for reparations. Yet moral indignation and resentment, so denuded of a desire even to have offenders suffer guilt, become objective attitudes. A totally benevolent being desires offenders to make reparations in the way you or I desire a rainstorm to cease. Blame is not involved in either case, but the offender is now treated as one would treat a nonagent. An agent almost always wishes to be treated as an agent more strongly that she wishes to avoid punishment.

The desirable trait of benevolence becomes undesirable when taken to too far an extreme. Total benevolence is thus a false ideal. We want to show goodwill to others, and should strive to avoid punishing wrongdoers more than is necessary for their reformation, or to repay society for their wrongs. Yet, punishing agents who do wrong shows more respect for them than does treating them as nonagents. One can love one’s neighbour and still wish her to feel guilt for the wrongs she has done. Therefore, it is not universal goodwill that people seek, but a wish to remember the humanity of everyone, including wrongdoers. This wish leads us to want to refrain from punishing anyone unduly, but not to the abolishment of a desire for offenders to suffer some form of punishment, even if it is only guilt. What we should count as due
punishment is the subject of much debate, and is beyond the scope of this thesis. But total benevolence is incompatible with our moral sensibilities, and is not the ideal for humanity that some think it to be. While a community should have compassion, it must also have a respect for persons.

In examining Watson's article, I have responded to some apparent examples of people who have given up the reactive attitudes, or some retributive elements associated with these attitudes. I have also shown how a definition of responsibility might clarify when the reactive attitudes are and are not appropriate. I believe that Watson has put forward the beginnings of a definition of responsibility, namely that moral understanding and some form of self-control are required of responsible beings. Jonathan Bennett also attempts to define responsibility for the Strawsonian model, and I will explore this attempt in the next chapter.
1. Strawson "Replies", 266.


3. Watson, 258.


5. Watson, 263.

6. Watson, 263.

7. Watson, 264.

8. Watson, 266.


10. Watson, 266.

11. Psychopathy is a matter of degree. I am here concerned with full-blown psychopaths. The following remarks will only apply to partial psychopaths insofar as they are psychopathic.

12. Watson, 277.

13. Robert D. Hare mentions such a test in Without Conscience: The Disturbing World of Psychopaths Among Us, (New York: Pocket Books, 1993), 34. The test is reliable, according to Hare, but requires the cooperation of the one being tested.

14. See Hare's Without Conscience for an excellent introduction into the nature of psychopaths and the history of how societies have tried to cope with them.

15. I refer to full-blown psychotics. To the extent that partially psychotic people are sane, we can use the reactive attitudes towards them.


29. Perhaps a totally benevolent being could also feel a form of guilt that includes the wish to make reparations for a wrong, without feeling that she should suffer for her actions. Guilt may not be possible without feeling bad, and hence suffering. Yet one can suffer without wishing further harm upon oneself. Thus a totally benevolent being might suffer guilt but only feel the wish to make reparations, not the wish to be punished.
Chapter VII: Bennett's Attempt to Define Responsibility

i) Minor Additions to Strawson's Theory

Unlike many other critics whom I examine, Jonathan Bennett agrees with most of Strawson's proposal. In "Accountability" Bennett points out some interesting facts about reactive attitudes.

First, he agrees that we can sometimes voluntarily hold the objective attitude towards normal people. This means that a person can, at least theoretically, choose not to blame someone for an action. This may be psychologically impossible, depending upon the potential blamer and upon the wrong that is done. But Bennett uses the possibility of withholding blame to help define blameworthiness. According to him, for a person to be blameworthy means only that it is not wrong to blame her, not that it is wrong not to blame her. Blame is not obligatory, but is permissible. Of course, as we can hold the objective attitude for only a short time towards normal people, we cannot stop blaming others for wrong actions in general. Nonetheless, since we are not obliged to feel reactive attitudes, they are more than mere recognitions of praiseworthiness or blameworthiness in others; they also express one's "emotional make-up".

Some people cannot officially withhold blame in this manner. Judges, for instance, must ensure that the law applies and is seen to apply equally. A judge's reasons are based upon an objective, instrumental model of what is best for society, however, rather than upon her reactive attitudes.
Yet it seems there are also cases in which someone is culpable if she fails to respond reactively towards a normal person. If a parent's young child is beaten by bullies, it may be a deficiency not to feel indignation on behalf of her child. This is because she fails to show the proper attitude of concern and protectiveness towards her child, with consequent effects on the latter's security needs. This may indicate that one can more justifiably forego resentment than moral indignation. Still, when one has no extraneous reasons that affect the suitability of resentment (or another reactive attitude) appropriateness alone does not entail that one should be resentful. There are many instances where one can properly choose not to blame another. As with hatred, sometimes it is psychologically healthier for a person to refrain from resenting another (to the extent that this is possible), even if she is entitled to her resentment.

Bennett also contributes to Strawson's argument when he lists new types of reactive attitudes. Bennett mentions that a person can be "morally indignant over someone else's attitudes to natural beauty". Although this indignation is more properly aesthetic than moral, it is a reactive attitude of sorts. This leads to new possibilities. Reactive attitudes do not have to be reactions to someone's attitudes towards a person, but could be reactions to her attitudes towards a work of art, or natural beauty, or non-agents such as the lower animals. This provides a possible basis for allowing animals some moral consideration, although I will not discuss in this thesis whether or not we are indignant at the
abuse of animals due to a violation of our moral values.

In addition, Bennett mentions a self-reactive attitude that is not a reaction towards what someone has done to others, but rather towards what she has done to herself. For instance, Bennett suggests that someone might reproach herself for having behaved foolishly. She may feel an intense form of shame, and she will consider herself responsible for what she did. Yet she might feel no guilt over what she has done, although she might if she also offended others.

These new examples cited by Bennett are at the edges of what might be considered reactive attitudes. They mainly fall outside the scope of morality. Yet, they are still concerned with responsibility. They have access to many of the same excusing conditions, and face the same prima facie threat from determinism. They are also supported against such a threat because of their being, like the other reactive attitudes, a natural part of our lives.

ii) The Difficulty of Defining the Reactive Attitudes

Bennett thinks it important to examine the reactive attitudes; a definition of these attitudes could lead to a definition of responsibility, or accountability. He holds that someone is accountable for an action "if a blame- or praise-related response to the action would not be inappropriate." This leaves open the question of what these blame- or praise-related responses, known as reactive attitudes, are. Therefore, Bennett
attempts to define reactive attitudes. While I hope to define responsibility by other means, it is still worth exploring how far Bennett progresses with his attempt.

Unfortunately, Bennett has difficulty defining the reactive attitudes. He first tries to see if attitude means manner of regarding, but fails to find any psychological or conceptual manner associated with, and only with, reactive attitudes. On the psychological side, Bennett's only candidate is to regard something heatedly, indicating a reactive attitude, or coolly, indicating an objective attitude. This form of coolness should not be confused with the cool reactive attitudes mentioned in earlier chapters. In any case, Bennett recognizes that psychological coolness can apply to reactive attitudes. One can have psychologically icy contempt. Additionally, one can have a hot objective attitude, such as burning curiosity. Thus, Bennett does not find the psychological manner of regarding to be a good definition of the reactive attitudes.  

As for a conceptual manner, Bennett conceives of two candidates. First, conceptual manner can merely involve bringing in the concepts of "person, action, and intention". Yet these concepts also belong to the objective attitude. For example, one may surmise that the reason someone runs from a snake is that she is scared of it, and, thus, wishes to escape.  

On the other hand, one can hold that conceptual manner means viewing someone as an unanalyzed whole. Still, this sort of viewing also can be found unaccompanied by reactive attitudes.
Bennett then investigates whether to hold a reactive attitude is to regard a person as a certain thing. This definition requires that reactive attitudes essentially involve having beliefs about another person. The beliefs would almost always have to be true since they supposedly define reactive attitudes. These beliefs would be facts pertaining to their objects. But with this account Bennett would have to define reactive attitudes only as facts, or at least what are thought to be facts, about other people. This definition ignores the "emotional" component of reactive attitudes that crucially distinguishes them from the objective attitude.¹⁰

More promisingly, Bennett tries to define a reactive attitude as "a posture of the mind which is apt for inter-action of a certain kind".¹¹ He thinks that he can define reactive attitudes if he can specify the kind of interpersonal relationship of which they are elements. This approach seems promising, but the problem is only pushed one step back. He must now find a definition now for the exact kind of interpersonal relationships for which reactive attitudes are appropriate. For instance, a psychiatrist and a patient have some sort of personal relationship, but the attitude involved in the treatment is not reactive. Bennett despairs of finding "some unarbitrary and uncircular constraint(s) on what is to count as a kind [of personal relationship]" for the purpose of defining reactive attitudes.¹²

Fortunately, we do not need a to find a definition for reactive attitudes in order to find a definition of
responsibility, although I defer the latter definition to chapter IX. Bennett successfully links reactive attitudes to interpersonal relationships of a certain kind. This proves useful for Bennett's later proposals.

iii) The Objective Attitude and Teleological Inquiry

Before looking at the link between reactive attitudes and interpersonal relationships, however, it is instructive to look at Bennett's characterization of the objective attitude. Bennett hopes that this will both shed light on reactive attitudes and explain the source and strength of the tension between the two types of attitude. He acknowledges that the attitudes can be mixed, such as when one is dealing with a growing child. Yet he maintains that a thoroughly objective attitude is incompatible with the reactive attitudes; one cannot hold the former and the latter simultaneously. Of course, contra Bennett, we may be unable to take a thoroughly objective attitude towards normal people; Strawson takes this view. Still, Bennett believes that one can temporarily cast out one's reactive attitudes by taking up an objective attitude. Certainly upon taking the latter one must almost entirely suppress the former. Thus, the tension between the two sorts of attitude is strong.

Bennett supposes that the source of the conflict might be due to "the limits on how much mental variety one can manage at a single time", or that "reactive attitudes essentially prepare for personal interaction while the objective attitude prepares for
inquiry, these two sorts of activity being somehow incompatible" (emphasis his). I believe that the tension might be conceptual. Just as one may have trouble conceiving a horse to be all white and all black at the same time, so one may have trouble conceiving a person to be both responsible and not responsible, or both an object to be handled and a person with whom to have a human relationship. In the case of a real horse, however, we can find out what colour it actually is, while in the case of our attitudes towards an agent, we can hold the objective attitude voluntarily towards an agent for whom we can also adopt the reactive attitudes. We can thus switch conceptions and temporarily choose to view someone as not responsible whom we would otherwise see as responsible. Therefore, the tension between attitudes may be more subtle than is that between conceptions. It is enough to accept that a strong tension exists between the reactive attitudes and the objective attitude.

Bennett defines the latter as an attitude of inquiry: "To adopt the objective attitude towards something is to inquire into how it is structured and/or how it functions." He is particularly concerned with those sorts of objective attitude that generate teleological inquiry, or inquiry towards an end, and not mere theoretical inquiry. He also holds that reactive attitudes never generate teleological inquiries, and proposes a property of reactive attitudes, if not a definition. "The attitude of x towards y is reactive if and only if it is a pro or con attitude which could not explain x’s engaging in teleological inquiry into
how y works"\textsuperscript{18} (emphasis Bennett'\textquoteleft s).

We will not find teleological inquiry in all cases pertaining to the objective attitude, as sometimes people adopt the attitude when they seek theoretical knowledge only. Nevertheless, Bennett believes that within the realm of pro and con attitudes, teleological inquiry is characteristic of the objective attitude.\textsuperscript{19}

Bennett suggests a third type of attitude which is neither reactive nor objective. "What I propose is this: if x's attitude to y does generate teleological inquiry, to that extent it is objective; if it could not do so, it is reactive; if it could but does not, it is neither"\textsuperscript{20} (emphasis Bennett'\textquoteleft s). As an example of this third type of attitude, Bennett proposes a passerby being disgusted by a beggar and hurrying away from her. Bennett does not think that the attitude of disgust is reactive, since "it is too close to what [one] might feel about a dead animal in the gutter".\textsuperscript{21} Bennett also maintains that it is not objective, since it apparently does not generate teleological inquiry, although, if the passerby had stayed to study the beggar in order to help her, then this inquiry could have been motivated by the passerby's disgust.

I disagree with Bennett on there being a third sort of attitude. I think that the attitude of disgust in the above example is objective. Furthermore, I believe that, although inquiry is often associated with the objective attitudes, we can also make teleological inquiries that have their source in the
reactive attitudes. For instance, I can be grateful to someone for doing a good deed for me and then try to discover what would be an appropriate gift for my benefactor. Bennett redefines what counts as a teleological inquiry in order to disallow this sort of case, but I believe that the distinction he makes is an ad hoc repair. He holds that teleological inquiry "seeks truths which can be applied in the pursuit of ends which are viewed as desirable either inherently or as means to some subsequent inherent good." Yet, my example seems to be a matter of teleological inquiry, contra Bennett. More importantly, I regard the objective attitude to be more essentially concerned with treating something as merely an object and not as a person, or subject, than with inquiries. I believe that the reactive and objective attitudes are complementary; I wish to avoid postulating a third type of attitude if an alternative explanation requires only two. I think that it is plausible to view the above case of disgust as an example of having an objective attitude towards someone, since the beggar is treated as an object. Therefore, I reject Bennett's proposal of a third sort of attitude.

iv) Interpersonal Relations and the Reactive Attitudes

Bennett is more successful when he examines the link between interpersonal relations and reactive attitudes. He notes that Strawson's account seems inadequate to explain the relation between reactive attitudes and some interpersonal relations. As Bennett sees it, on Strawson's original account one must have
interpersonal relations before one can have reactive attitudes. \(^{24}\)

Unfortunately, this is insufficient to explain the reactive attitudes one can feel towards a stranger who does something minor to one, good or bad. The action is immediate, and does not require a previous interpersonal relationship in order to be meaningful. The interpersonal relationship the action generates, if any, appears to be too meagre to give rise to reactive attitudes; it is not what we mean by the web of interpersonal relations of which people are a part, and which we would regret losing, if reactive attitudes were absent from humanity.

Bennett surmises that interpersonal relationships have a slightly different connection to reactive attitudes than Strawson outlines: "What should be emphasized, I suggest, is not the relations within which reactive attitudes arise, but rather the relations towards which they point"\(^{25}\) (emphasis Bennett's). Thus, one uses the reactive attitudes to create or attempt to create an interpersonal relationship, if it does not exist already between the person who holds the reactive attitude and the one who is its object. As a caveat, I believe that a person must be part of a network of interpersonal relationships before such reactive attitudes can be learned. Thus, a child raised in the wild cannot immediately have reactive attitudes merely by trying to create interpersonal relationships. She must rely on others to create interpersonal relationships with her first, and, thus, give her a sense of being part of a moral community. In general, however, I believe that Bennett gives a good account of how our reactive
attitudes work.

In some cases, a person may fail to create an interpersonal relationship, as when someone initially resents being pushed, but then realizes that the person doing the pushing is having a psychotic episode or is, in some other way, excused from being responsible for what appears to be an expression of a harmful attitude. Bennett’s explanation of the reactive attitudes thus seems to fit in with Strawson’s account of excuses, and proves useful in resolving some puzzles.

We find some of these puzzles when we examine the self-reactive and vicarious reactive attitudes. In both cases, especially the latter, it is hard to see exactly how the interpersonal relations are set up. If Arnold is indignant because of Betty’s treatment of Carol, then who has the interpersonal relationship that leads to the moral indignation, Arnold and Betty or Betty and Carol? Yet, if the reactive attitudes create an interpersonal relationship, we can answer this question. Arnold is indignant, and, therefore, actually or imaginatively creates a relationship between himself and Betty, the person he blames. He thus makes what Betty does to Carol his business, even if he had never before had anything to do with either woman. In the self-reactive case, one imaginatively forms a relationship between one’s present self and a past self. This way of understanding reactive attitudes as creating the grounds for addressing the wrongdoer (or rightdoer, as the case may be) is adopted by Watson in "Responsibility and the Limits of Evil."
He sees Bennett's account as a useful way of looking at reactive attitudes. Watson uses it to examine the underlying grounds for the effectiveness of the excuses that disengage the reactive attitudes. Thus Bennett, with his account of the link between reactive attitudes and interpersonal relationships, may indirectly help to form a definition of responsibility.

Bennett's account works for the positive reactive attitudes, such as pride, gratitude, and moral appreciation, as well as the negative ones. In each case, one can think of these reactive attitudes as setting the stage for an interpersonal relationship, if one does not already exist. It does Bennett credit that his account can adequately interpret both the positive and the negative reactive attitudes.

Bennett provides a helpful examination of the theoretically nonimperative nature of blame, and of various types of reactive attitudes not considered by Strawson. He also presents a useful way of looking at the link between reactive attitudes and interpersonal relations. Nevertheless, he fails to find a definition for the reactive attitudes. As the reactive attitudes seems to be linked to responsibility in some way, this indicates that it will be difficult to formulate its definition.
5. Strawson notes that there are some behaviours that are not attitudes towards people and yet may receive reactive attitudes as a response in his "Replies", 266.


7. Bennett, 32.


11. Bennett, 34.

12. Bennett, 36.


15. Bennett, 40.


17. Some cases of reactive attitudes spark theoretical inquiry. For instance, one could be grateful for an unexpected inheritance, and through inquiry, find out that one received it due to a resemblance to the benefactor's late husband. Earl Winkler provided this example.

18. Bennett, 38.


20. Bennett, 38.

21. Bennett, "Accountability", 38. Note that this form of disgust is different from the moral disgust we would feel towards a moral reprobate. It is unfortunate that the same word describes two separate emotions.
22. Bennett, 37.

23. Bennett, 37. One could argue that helping one’s benefactors, when feasible, might be an inherent good, though a supererogatory one. This is not to say that it could not conflict with other inherent goods.

24. Bennett, 42.

25. Bennett, 43.

26. As described below, the relationship that is created can be hypothetical. If one later finds out that people laughed at one’s being pushed, then resentment would create a hypothetical relationship with them, even if one never met them.

27. Bennett, 44.


29. Interestingly, reactive attitudes, or a lack thereof, can themselves be reacted to. A lack of gratitude on one person’s part may lead to resentment on another’s.
Chapter VIII: On "Moral Luck"

i) Variable Consequences and Moral Tests

I wish to examine the intuitions behind the idea that it would be irrational to praise or blame people if determinism were true. I will not argue against these intuitions here, as I believe that Strawson and Wolf have adequately shown that it is not irrational to hold the reactive attitudes, regardless of the validity of determinism. By inspecting these intuitions in detail, however, I may be able to find out why they seem attractive. I also hope to explain why certain things should be excluded from a definition of responsibility.

The intuitions in question are clearly expressed in Thomas Nagel's "Moral Luck". Nagel divides them into four basic types, each of which seems persuasive. Together, they appear to leave the moral self with nothing to do, and, in a sense, nothing to be.

The first category of moral luck deals with how the results of one's plans are affected by external circumstances. Nagel examines the case of negligence, exemplified by a drunken driver who swerves on to a sidewalk. We consider him a more blameworthy agent if he hits and kills a child than if he merely drives on to an empty sidewalk. Yet, whether or not the child is there is beyond the driver's control. On the heroic side, we praise a fireman who saves a child from a burning building more than we do one who tries but fails.

In addition to cases like these, Nagel looks at decisions made under uncertainty. "Someone who launches a violent
revolution against an authoritarian regime knows that if he fails he will be responsible for much suffering that is in vain, but if he succeeds he will be justified by the outcome." Also, people can know in advance that whether they will be held mildly or greatly culpable for an action, or even whether they will be praised or blamed for it, will depend on many external factors beyond their control.

Moral agents are subject to variations of blame and praise based upon the results of actions, many aspects of which are beyond an agent’s control. Although this is an interesting fact about moral responsibility that deserves more discussion, it is not relevant to my thesis. Unlike some other forms of moral luck that Nagel discusses, this one does not support an intuition that people are not responsible agents at all. Whether a drunken driver is responsible for manslaughter or not, he is still responsible for driving while impaired.

On the other hand, this type of moral luck illustrates how we treat responsible agents differently, depending upon the outcome of their plans and desires. This seems to be unfair. We judge the intention of an agent who attempts an action or makes a decision under uncertainty; yet, if the consequences of her action are variable and can be affected by external factors beyond her control, why do we also judge her for these consequences? This sort of judgement is apparently part of human nature; people blame a criminal less for an attempted murder than for a successful one. Yet is human judgement rational in these cases?
Wolf's and Strawson's defences of the rationality of taking reactive attitudes do not settle this question, since we continue to hold the actor to be a responsible agent. Thus, the question of whether we would be more rational if we did not judge results, but only intentions, is an open one.

It is possible to point to some considerations that support the rationality of judging people on the basis, not only of their attitudes, but also the actions resulting from them. First, there are legal problems, such as how to punish the drunken driver in the above case. Should we let someone guilty of manslaughter off with a fine? Surely not. Yet it also seems impractical and unjust to give everyone found guilty of drunken driving a sentence equivalent to that for manslaughter.4

Practical matters aside, people apparently must accept that each action they choose contains an element of risk. Thus, a person who chooses to drive while drunk runs the risk of getting caught, or of seriously harming herself or others, but decides to take that risk. Therefore, if she does get apprehended for impaired driving, she simply gambled and lost, and, thus, has no reason not to be punished for her action.5 If she kills someone while driving impaired, then she has also lost her gamble, and can be blamed for taking such a huge risk with other people's lives. Yet, if she does not kill someone, then, although the initial risk is the same, she has won the gamble. She is still to blame for drunken driving, but is not guilty of manslaughter.

Daniel Dennett employs a sports metaphor to portray a
counterintuition to the apparent unfairness of blaming someone for the results of her actions. In basketball, we do not discount a shot merely because it was lucky, or allow a shot that barely missed to count for points. Yet, external factors beyond the control of the shooter may have affected the shot, such as how she caught the ball just before she shot, or that a key defensive player was in an accident and so could not play in that game. Luck is simply part of the game, as it is a part of life; we do not consider disallowing lucky shots in the game. Should we disallow unlucky results in life? Perhaps, if the results were unforeseeable, but one of a number of foreseeable possible outcomes occurs, then it seems fair to hold someone responsible for choosing a path that could lead to that outcome, among others, even if the person does not want that particular outcome.

Dennett believes that, given that we are finite beings who have to make decisions within a limited period of time, we are entitled to be somewhat arbitrary in our manner of holding others responsible, in order to promote wiser risk-taking strategies, usually strategies that minimize the risk to others. This provides an objective reason to take certain reactive attitudes, which appears to undermine Strawson's account of the framework of the reactive attitudes as not needing a defence against external criticism. Yet, Strawson does hold that, given that we are responsible beings, utilitarian considerations in moral condemnation and legal punishment still have a role to play. In any case, our concern to reduce dangerous risk-taking strategies
does not invalidate the reactive attitudes that motivate our punishing people for the harmful results of such risk-taking. Therefore, blaming a person for the results of her actions, as well as for her attitudes towards other people as expressed in those actions, appears compatible with Strawson's viewpoint. Nevertheless, this compatibility rests upon our refraining from blaming people merely for the results of their actions. We must also take into account their intentions. Thus we excuse agents from responsibility for unforeseen accidents. We do not blame someone who pushes another in a crowd solely as a result of being suddenly jostled herself.

This points to one factor that should not be in a definition of responsibility. We should hold people responsible for foreseeable effects of their choices. Yet we should not require that only people able to predict the exact results of their actions are to be considered responsible agents. This will modify some views of responsibility. We thus sometimes hold people responsible for the results of their actions, even when they did not desire those particular results. We hold them responsible so long as those results were foreseeable as possible outcomes of her actions. Therefore, to be responsible, an agent requires only a general knowledge of the possible foreseeable results of her choices.

By investigating the above form of moral luck we discovered some important facts about the nature of responsibility and of our reactive attitudes. The next form of moral luck seems to
undermine the rationality of reactive attitudes more seriously. This type of moral luck is exemplified by the Nazi concentration camp guard who would have led a quiet life in Argentina if he had left Germany for business reasons in 1930, or by the citizen of Argentina who actually led a quiet life but would have become a Nazi guard if he had left for Germany in 1930. There are certain moral tests that some people face. They pass or fail these tests, and we praise or blame them accordingly. On the other hand, there are other people who never face these moral tests. If they had, then they would be candidates for praise or blame for how they reacted. As it stands, although they are still moral agents, they have a different moral record because they have not faced that particular moral test. Furthermore, it seems a matter of luck whether they have to face that test within their lifetimes or not.  

This form of luck, like the first one, does not deny that there are responsible moral agents. It does show that blaming or praising an agent depends, at least partly, upon tests that she either does or does not face and whether or not she faces them is controlled by factors external to her. "A person can only be responsible for what he does; but what he does results from a great deal that he does not do".

I think that a defence can be made of the rationality of our reactive attitudes in this case. Although our reactive attitudes are primarily reactions to others' attitudes, we cannot always easily determine what those attitudes are. When someone faces a
moral test and acts well or badly, we can examine her action and conclude what her attitude must be. But when a person does not face such a moral test, we have no means of knowing what her attitude would have been. Since we do not have enough data to judge what her attitude would be in such a situation, we do not see her as especially heroic or villainous. Rather, we render a fairly neutral judgement of her, at least in respect to tests that she has not faced. It is not irrational for us to withhold judgements when ignorant of data essential to making those judgements. Once a person passes or fails a moral test, then our reactive attitudes come into play.

This reveals another element to omit from a definition of responsibility. We should not require an agent to know exactly what moral tests she will face. Certainly she should not have to face a trying moral test in order for us to consider her a responsible agent.

ii) Character, Attitudes and Contra-Causal Freedom

Although we cannot know what tests we will face, we can, in general, prepare to meet them by trying to be good people, that is, to have the proper attitudes towards others. This inculcation of good character will help us pass the moral tests we might face. On the other hand, there are those who argue that nobody is responsible for her own character. These people argue that we are victims of moral luck in character formation. This third sort of moral luck, unlike the previous sorts, is an attack on a
person's being a responsible agent. We have seen one example of it in the case of the murderer Robert Harris, as described by Watson in chapter VI. Harris's character seems to have been formed through his being abused by his parents and by his fellow inmates. Furthermore, we, as observers of Harris's history, cannot be entirely certain that we would not have become as cruel, vicious and cold-blooded as he, if we had had the same upbringing. "The fact that my potential for evil has not been nearly so fully actualized is, for all I know, something for which I cannot take credit...only those who have survived circumstances such as those that ravaged Harris are in a good position to know what they would have done."12 This example combines two forms of moral luck: that involved with having to face a moral test and that involved in the formation of one's character. The latter is more appropriate, likely to make one wonder about Harris's bad luck, however, since his unfortunate formative circumstances began when he was very young, before he could have been considered a moral agent at all.

My answer to the problem of luck affecting the formation of one's character is similar to the answer to the problem of luck involved in what tests one will face in life. We do not know how we would have acted had we had the upbringing Harris had. We know that some people with just as horrible an upbringing grow up to be moral adults. We also know that Harris turned out to be evil, and we can judge him for his evil actions because of this.

Ferdinand Schoeman similarly points out that a disproportionate number of people who grow up in poorer
neighbourhoods will turn to criminal activity. An accident of their place of birth seems to have been a factor in their becoming criminals. The same reply can be made in this case as to the case of Robert Harris. We deplore living conditions that can influence one to become a criminal, and as a society we should do everything we can to provide people with decent environments to grow up in. Still, we hold people responsible for their actions, whatever upbringing they may have had. We know that some people transcend unfortunate backgrounds and become good citizens. We hold the same requirement of character formation and self-evaluation for people with these backgrounds as for people with more fortunate ones. There are certain attitudes towards others, such as a relatively stable benevolence, that we as members of a moral community simply require of each other. Thus, although as a society we are obliged to impart a moral education to our citizens, they are still faced with the requirement to be good on their own. Therefore, we blame them for wrong actions that result from malevolent attitudes.

We also judge people for these attitudes themselves, even if the attitudes are not manifested in actions. Nagel gives the example of the man who cannot help feeling envious of others, although he does not reveal this to others or do anything to impugn their success.

Whether or not it is irrational to judge someone for attitudes that she may not have had any control in developing is, I think, answered by the following. We are all, to some extent,
subject to the whims of fate in the attitudes that are impressed upon us when we are very young. Yet humans beings can and do reflect upon the appropriateness of their own attitudes. Thus, we are capable of modifying our attitudes to a great degree. For example, a child growing up in a racist household both can and is expected to reflect upon her racist attitudes and thus overcome her racism, even if this is more difficult for her than for someone who grew up in a more liberal household. By the time we are adults, we should be able to judge those attitudes we hold that are morally repugnant and work to reduce them. Thus, we are responsible agents whom people count on to be good and to show the proper attitudes towards others. Charles Taylor makes the point that, once we are adults, our character formation is our responsibility. "Because this self-resolution is something we do, when we do it, we can be called responsible for ourselves; and because it is within limits always up to us to do it, even when we don’t...we can be called responsible...for ourselves whether we undertake this radical evaluation or not." Dennett makes a similar point in chapter 4 of Elbow Room. Therefore, another fact about a definition of responsibility is that it should not require us to meet the impossible condition of creating ourselves from scratch. Nevertheless, it may require us, to a considerable degree, to be able to modify ourselves progressively to become agents of better and better moral character.

Harry Frankfurt raises another issue concerning the reformation of one’s character. He stresses that people are
responsible only for those actions that are willed by them. Furthermore, they are responsible only for the will that moves them to actions, if it in turn is a will that they desire. They must identify with the will as "their own"; they must not have the will that moves them to action be a will that they disapprove of, from a higher level of judgement. Frankfurt gives the example of an unwilling drug addict, whose desire to take the drug will cause her to take it. She wishes that this desire did not move her to action, but, rather, that her desire to be free of the drug (which she identifies as more her own desire) prevailed and caused her to refrain from taking the drug.

I sympathize with the view that addictions of this sort can be treated as a sickness and, thus, as out of the control of the agent. Nevertheless, people who commit antisocial actions, such as stealing or killing, in order to feed their drug habits, are not, thereby, either legally or morally excused from responsibility for stealing or killing, regardless of whether they identify with their desire to feed their addiction or not.

It may be that a person is now so far gone in her addiction that she can no longer stop herself from taking the drug, even if she underwent the best drug rehabilitation treatments. Perhaps then we would not consider her responsible for taking the drug. Yet, if this is so, I do not see how she would become more or less responsible for drug-taking merely by being happy or unhappy about being a drug addict. Her higher-level desire not to want the drug might show that her wish to reform her habits is a sign of good
character. Still, this does not make her less responsible for taking the drug. Therefore, I think that an agent's wish to distance herself from certain of her desires makes her no less responsible for the actions that result from those desires, although her wish to reform her moral character is itself praiseworthy, and, thus, makes her overall a better person than someone who does not try to reform her character. I think that a significant point we can glean from Frankfurt is that moral agents, and only moral agents, are capable of evaluating their own moral characters in this way, whether they do so or not.

Yet, if determinism is true, then some people are determined to succeed in the reformation of their characters, whereas others are determined to fail. Thus we turn to the fourth type of moral luck, which is the original doctrine of determinism raised in chapter II. Of the four types of moral luck, it seems to be the most threatening to people's sense of being responsible agents. I have already addressed the concerns raised by this form of moral luck, however, and I will not recount them here. I will simply note that a definition of responsibility cannot require one to have contra-causal freedom, since determinism neither can nor should affect our holding people responsible.

iii) Christian-Kantian Thought

Although the four forms of moral luck that Nagel describes have been shown not to be a threat to an agent's responsibility, they certainly seem intuitively convincing upon first glance. Why
does the doctrine of moral luck so easily influence us to believe that nobody is responsible? Jonathan Bennett proposes one fairly convincing answer.

Bennett describes the Spinozistic view, which is the attempt to justify taking the objective attitude towards everyone all of the time. He states that the Spinozistic view draws support from what he calls Christian-Kantian thought. This thought maintains that moral worth cannot depend upon natural advantage. The basis of this view is that God or Nature created people with the ability to transcend their natural causal history; thus, we can choose to be good or evil, regardless of their empirical conditioning. Otherwise, it seems unfair for a God to create our natural circumstances, create us within the context of the natural realm, and then blame us when we act as we are empirically conditioned to act.¹⁸

Christian-Kantian thought has a lot of support in our culture, but it requires of each of us the seemingly impossible condition of being able to create from scratch one's entire moral character. As Bernard Williams says, "No human characteristic which is relevant to degrees of moral esteem can escape being an empirical characteristic, subject to empirical conditions, psychological history, and individual variation".¹⁹ Bennett, 26.²⁰

The popularity of Christian-Kantian thought has consequences. People are capable of suspending the reactive attitudes towards responsible agents for a short time. This capability, combined
with how Christian-Kantian thought makes people feel that they, in some sense, should exercise it, leads them to believe that Spinozism is rationally required. Bennett characterizes their situation as follows: the temporary taking of the objective attitude leads to a momentary evaporation of the reactive emotions. People then compare this evaporation with other situations in which their emotions have been disengaged, such as when they were at first afraid of something harmless, but then saw that there was no reason to fear. They may then falsely conclude, because for a brief time they no longer feel the reactive attitudes, that Christian-Kantian thought has provided a counter-argument to the rationality of these attitudes. Furthermore, they are not, in general, used to seeing blame and praise as reactive emotional attitudes so much as intellectual operations. Most people are not Strawsonian theorists, and

"When they are swayed by Spinozism don’t think of blameworthiness, etc., in terms of feelings at all, and so they will tend to construe their loss of indignation, etc., as the loss of the judgement that the subject is "to blame"; and when a judgment is sent packing by an intellectual operation one naturally thinks that it has been refuted or at least discredited"21 (emphasis his).

Thus, people think that the reactive attitudes are disqualified when they are really only temporarily dispelled.22 This may help to explain why Nagel’s examples of moral luck seem at first to be so persuasive.

Overall, Nagel’s article provides some insights into the nature of our reactive attitudes. Although I think that the
defence of our being, on the whole, responsible creatures can withstand his assault, he does show how complex our reactive attitudes can be. For instance, we take the externally conditioned results of actions into account when judging someone’s attitude towards others. Also, by examining these forms of moral luck we have uncovered some things that are not part of a definition of responsibility. We should not require human beings to do any of the following: 1) be able to predict the results of their actions exactly, 2) know what moral tests they will face, 3) be able to create their own characters from nothing (although they should be able to modify the character they happen to have), and 4) have contra-causal freedom. Still, these negative conditions leave open the question of what does go into a definition of responsibility. It is to this that I now turn.

2. Nagel, 175, 178.

3. Nagel, 179.


5. See Dennett, Elbow Room, 162, for a further exploration of this aspect of moral luck.

6. Dennett, 94.

7. Dennett, 164-5.


9. Nagel, 175, 182.


11. Nagel, 181. This form of luck is actually mentioned before that of the moral tests a person faces, but I felt it more appropriate to alter their places in order to show the types of moral luck in increasing order of the scope of their attacks upon a person's sense of being a responsible agent.


17. Frankfurt, Harry G., "Freedom of the Will and the Concept of a Person" in Free Will, 81-95.

18. In fact, Bennett holds that Christian-Kantian thought is logically unsatisfiable, since it requires us to be "free" or "autonomous" or "self-made" in a sense in which it is logically impossible to be free or autonomous or self-made." Bennett, 24-5.

19.


Chapter IX: Responsibility

i) Reactive Attitudes and Reflection

Any definition of responsibility congruent with Strawson's views should take into account our reactive attitudes. As we cannot dispose of these attitudes, it is instructive to examine what they cause us to do. For example, one of the major effects of guilt is to persuade us to make reparations for our misdeeds. More generally, the reactive attitudes dispose us to reform our character, or that of others.

Reactive attitudes thus motivate us to improve our character. We cannot help having this motive, since we cannot help being creatures with reactive attitudes; yet, we also have moral reasons to improve our own and others' character. Although they differ on many points, most moral theories agree that it is generally a good thing to improve one's own moral character, or to assist in improving that of others. Thus, the reactive attitudes influence us to do something that is considered morally good.

Another morally significant natural fact about us is that we are capable of reflecting upon our own dispositions and character - how we tend to act in certain situations, or which risk-taking strategies we take. An example of the latter is the degree of attention we pay to the road while driving, usually less on an empty highway than on a crowded city street. Moreover, this capacity for self-reflection is very highly developed. We can represent our reflections to ourselves using language, either in speech or in thought. Our self-reflective ability has been
examined by Taylor, but is more closely scrutinized in Dennett's Elbow Room. Dennett points out that our brains approximate the behaviour of perfect rational wills. Although we are not perfectly rational creatures, taking every significant modification of our environment into account, our brains "are cunningly designed to be not infinitely but indefinitely sensitive to meaningful changes." We can correct our occasional failure to take significant facts into account through further self-reflection. Unfortunately, we cannot continue the process of reflection forever - eventually we have to act - yet, we can continue indefinitely. We cut off the reflection process when we feel that we have thought about something enough.

Our capacity for self-reflection is like the ability of birds to fly. Neither is negated by determinism. Furthermore, the general ability to reflect upon our own character, and upon the actions which tend to spring from it, is what allows us to be responsible agents. When this ability is absent, so is moral responsibility. For various reasons, psychopaths, children, and psychotics do not (fully) have this ability, and so are not (fully) responsible. Recognizing this, we do not hold them (fully) responsible for their misdeeds.

Of course, we are subject to factual errors about who has the ability of self-reflection. In the past women, and people of certain races, were sometimes considered not (fully) capable of self-reflection, nor thus of moral improvement. They were treated as less than fully responsible beings. On the other hand, some
lunatics were thought more capable of reflection than they were, and thus were treated as more responsible. When we discover such factual errors, we correct them and proceed to feel reactive attitudes towards these people to a more appropriate degree. Most adults are considered moral agents, responsible for their character and through it their actions. We have the capacity to improve our character through self-reflection, and to have this capacity is to be a responsible being. Thus, since most humans have such a capacity, they are responsible beings.

ii) The Basis of our Excuses

Self-reflection is not the only general capacity that humans have. Most also have the general capacity to do a number of tasks, from kicking a soccer ball to cooking a five-course meal. This is because humans have the general capacity to learn, and unless they are physically or mentally incapacitated, can learn to do a wide variety of physical and mental tasks.

People also have the general capacity to be good. That is, most people can learn the minimum moral requirements put on them by society, and can meet those requirements, in the sense that anyone with at least one arm can throw a baseball.

This general sense of ability is not what we usually mean when we say that someone is able, for instance, to play hockey. Here we think of the hockey player as having mastered a number of skills. When we say that she is able to score a goal when she is 20 metres from the net we mean that she succeeds in scoring goals.
at that distance, either all of the time, or most of the time, or significantly more often than someone untrained in the skill. We do not say that she has the ability to score a 20 metre goal if she succeeds only one time in a thousand. We require of people a level of success, varying with each skill, before we will say that they have mastered that skill, and are able to do the task associated with it.

People can improve their skills by inculcating good habits and removing bad ones. Our hockey player may have learned to grip her stick correctly and not to blink when she shoots. While there is always room for improvement, a person will reach a point when she has learned enough good habits and eliminated enough bad ones that we consider her able to do a task, such as play professional hockey. To distinguish this sense of ability from the sense of general capacity outlined earlier, I will refer to the latter as a general capacity, and the former as performance-ability.

There is a third sense of ability, which I will call micro-ability, which might be used to describe whether a person, given her exact physical micro-state and that of her environment at time t, will succeed or fail at task x. We do not use probability to describe a person's micro-ability to do task x, since if she succeeds her chance of success is 1, and if she fails her chance of success is 0. Incompatibilists seem to focus on micro-abilities. They think that talk of performance-abilities ignores too many environmental factors to be relevant to a discussion of responsibility. Conversely, micro-abilities seem to be useless
when describing human performance-abilities. The person and the environment will never repeat their exact physical micro-state. Furthermore, we will never even be able to describe such a micro-state well enough always to distinguish it from those of the same person and her environment at a different time $t'$ when her attempt has a different outcome than at time $t$. In fact, the attempt at time $t$ might be the one time in a thousand that an unskilled person succeeds at task $x$, or that a skilled person fails. Thus, we cannot use micro-abilities to describe whether a person is skilled at a task or not.

Moral responsibility is linked to the first two sorts of abilities. Most adults have the general capacity to be good. They are aware of what will help or harm others, and they are physically capable of being helpful. In addition, due to their general capacity of self-reflection, most people have the performance-ability to be good. Unlike with other performance-abilities, people are expected to exercise the performance-ability to be good at all times. Failing to exercise this performance-ability thus is itself a sign of a low performance-ability to be good. Sometimes one's performance-ability falls below a generally accepted level. This is because a person has not acquired enough good habits of character, and has retained too many bad ones. Thus, they are not proficient enough at being good to be considered people of good character.

Here too, we do not consider micro-abilities when we judge a person's moral worth. We consider some risk-taking strategies
to be morally acceptable, even if they do not pay off in a particular instance, and others to be unacceptable, even if they do. If a careful driver hits a child who runs in front of her, then, if she took every reasonable precaution, she is still allowed to drive. A drunken driver, even if she hits nobody, is still considered unsafe.

It is important to note how much emphasis we put on our moral habits and how they affect our performance—ability to be good. They form a part of our character, which our reactive attitudes influence us to improve. Sometimes we cannot, as far as we can see, significantly improve a habit or strategy of action to avoid tragedies such as the careful driver’s, described above. In these cases we hold nobody responsible for bad results; they are written off as accidents. The inability to improve one’s habits (not because of some incapability but because there are none better to be found) allows us to excuse accidents, some cases of ignorance and some forms of coercion. We also excuse events that the agent cannot avert, regardless of which strategy she employs, or what habits she has. For example, a police officer rushes to save a baby in a basket hanging by a fraying rope over a one hundred metre drop. The rope breaks before she can get to it, as she knew that it almost certainly would. She should not be considered responsible for the baby’s death, because she could have done nothing to save the baby. Indeed, the officer tried the best strategy available, even though its chance of success was minimal. In general, a person is not considered responsible for accidents,
etc., when her character cannot be improved upon to enhance her chances of avoiding similar unfortunate happenstances in the future, or at least not without disproportionate sacrifice of other important dispositions and goods.

Yet we do not excuse all cases of ignorance and coercion. We expect a person to be reasonably aware of her surroundings and to take suitable precautions. To avoid unnecessary ignorance, she should gather enough information to act appropriately in most situations. We expect a certain level of knowledge of her. This level may vary with the role she assumes. For instance, we expect a doctor to know more about what to do if someone is choking than is a layman, and as a doctor she will more likely be held responsible if the choker dies because of her ignorance. Similarly, we expect a person to avoid situations in which she will fairly likely be coerced; we demand that she take reasonable (though not excessive) precautions. Moreover, we also hold her responsible if the coercion is minor and her action very harmful. In these instances, a person can modify her character, including her information-gathering strategies, to become better able to deal with the world, and less likely to be a threat to herself or others. That is why we hold her responsible.

Character improvement may also explain why we excuse cases of temporary stress. During these periods, a person abandons her fine-tuned character traits and risk-taking strategies for cruder ones, because more of her brain is occupied with dealing with the stress. This may explain why people under stress often make
stupid mistakes. The ability to reflect is one of the first fine-tuned abilities that a stressed person abandons. Thus we wait for the short period of stress to pass before we consider her to be once more responsible. If the periods of stress are longer or more frequent, then we blame the person for not developing a character in her less troubled periods that can better deal with stress, and thus sooner re-engage her more competent character traits. On the other hand, some people have gone insane after suffering severe disasters. This shows that, if the periods of stress are severe enough and long enough, we recognize that the person is insane. She is no longer a responsible being, since she ceases to have the capacity of self-reflection at all.

The manner in which we say that people are responsible for specific actions may seem arbitrary, especially since it is motivated by reactive attitudes that we cannot help but feel. It makes sense, given reactive attitudes, not to think of responsibility for an action as something we simply happen to have. Rather, we should think of responsibility as something we either take for ourselves or impose upon others, for the purpose of improving character. We thus see taking and imposing responsibility as natural facts of our behaviour: they are the expressions of our reactive attitudes. It does not seem irrational to take or ascribe responsibility, since our desire to improve character is based upon our reactive attitudes, and we have shown in chapter IV that they are not irrational to hold. Moreover, since it is considered a morally good thing for a person
to improve her own or another's character, we should consider taking and imposing responsibility also to be morally good. Of course, this applies only when to take or impose responsibility will improve a person's character so as to decrease the chance of her performing similar wrong actions in the future.

iii) Why We Want Responsibility

Why do we want responsibility? Because we have reactive attitudes, particularly self-reactive attitudes, which make us want to improve our character. By taking responsibility for our actions, we acknowledge that one of our actions, or thoughts, is bad and shows us that we have a character flaw within ourselves. We also pledge to strive to improve our character so that, in the future, we will not perform relevantly (or sometimes roughly) similar actions in relevantly (or sometimes roughly) similar situations. This only makes sense for actions that we feel occurred because of faulty character. If we tried to take responsibility for other events, such as a lunar eclipse, there would be no particularly relevant character flaw to correct. We take responsibility only for events in which our risk-taking and information-gathering strategies, our values, our habits - in short, our character - play a role. Further, we take responsibility for such events only if the role that our character plays is a general one, likely to be repeated. Although I may, in trying to help someone, inadvertently harm her due to unforeseeable circumstances, it is still a good character trait
to help people in general; we may not need any self-modification.

As for us imposing responsibility upon others, I believe that Watson's description of reactive attitudes as moral address, summarized in chapter VI, is significant. The imposition of responsibility upon others is the process of our addressing them, and demanding that they reform their character. Sometimes our moral address is hypothetical; we may become indignant towards a dictator whom we will never meet. In these cases it may seem that our reactive attitudes do not fulfil their function of attempting to improve another's moral character. Yet even here we may with our reactive attitudes still have a beneficial effect upon people whom we do encounter. We all act as role-models for one another, and by feeling and expressing our moral indignation at an absent party's evil actions we let those around us know that we consider such actions reprehensible, and a sign of bad character. Thus, we are not irrational to ascribe responsibility to people whom we will never meet.

Responsibility for actions reduces to responsibility for a character that results in such actions. It would be better if we had a method of judging someone's character directly, perhaps by reading their minds, but alas, we are not telepathic, and the best way we can judge someone's character is through their actions.

iv) Minor Objections

Some people may feel that my definition of responsibility is inadequate. For instance, I have concentrated upon negative
reactive attitudes, such as guilt, resentment, and moral indignation. What about positive reactive attitudes, such as gratitude and moral praise? They certainly cannot act to correct a defect in character: do they have another purpose? I think that the positive reactive attitudes serve to reinforce a good character, one that tends to result in proper actions. These are neither as strong nor as well-defined as negative reactive attitudes. We usually do not praise people merely for doing what is proper, since that is expected of them. We reserve positive reactive attitudes for those instances in which people act above and beyond the call of duty, or at least beyond our expectations. Positive reactive attitudes may not require us to be saints, but they do work to reinforce saintly behaviour should it occur. We are usually more concerned with negative reactive attitudes; we worry more about the dangers of bad character traits than of merely mediocre ones. Yet the positive reactive attitudes still play a role. Sometimes people act better than we expect that they should; sometimes they act worse. The reactive attitudes influence us to avoid bad character traits while maintaining good ones.

Some people also object to the idea that all of one's actions are a result of her character. They might argue that I am ignoring spontaneous acts, such as tossing someone off a cliff on a momentary whim. I think such acts, when they are harmful, can still be linked to a character flaw. In the above case, this flaw is to behave with too much spontaneity in an inappropriate
situation. Nobody should be the sort of person who throws people off cliffs on whims, or harms people on whims in roughly similar situations, such as running them over in one's car.

v) On Freedom

Finally, people might object that, although I discuss moral responsibility and reactive attitudes, I overlook the concept of freedom. I focus mainly on responsibility, not freedom, because I believe that the age-old problem of freedom and determinism is mainly about whether or not we can be morally responsible in a determined world. Nevertheless, I will say a few words about what freedom is, and whether it is compatible with determinism.

I have mentioned that people have the ability to do certain tasks. In life, we perceive opportunities to bring our abilities into play, to have an effect upon our environment, or at least, as far as we can see, to increase the chance of doing so. We feel that we can act to make a difference to what happens, preferably a beneficial one. The awareness of these chances to use our abilities is our freedom of perceived opportunity. For example, a hockey player might see an opportunity to move closer to the net in order to exercise her ability to score a one meter goal.

On the other hand, she might miss an opportunity to score if she does not notice it. A spectator might notice such a missed opportunity, and she might be able to show the player that she missed a chance to score.

A person might also believe that she sees an opportunity
where there are none. For instance, she might buy a ticket from a crooked lottery operation. She will believe that she has an opportunity to win, but all she has is a false opportunity.

How can we tell when an opportunity is real? Ultimately we cannot, since that would require an understanding of the microstates of our environment. While we can usually apprehend most environmental factors that provide or deny opportunities to act, we cannot possibly notice all of them. The best we can do is first, to strive to become more aware of our environment so that we can perceive as many opportunities to use our performance-abilities as possible, and second, to inform each other of relevant factors that the other person does not notice. Still, some opportunities will be missed. If nobody can see, for instance, any opening for a hockey player to score a goal (the defence might be too good) then we would say that she has no opportunity. Perhaps by some strange combination of events she might have scored. She might have hit the puck at the ceiling of the stadium, at which point it would have rebounded off of a thrown beer can to fly into the net just when the goalie accidentally trips. Nobody would suspect this opportunity, since nobody could have foreseen it. Thus the freedom of perceived opportunity, or freedom of opportunity for short, is subjective, or at least intersubjective, and we can increase this freedom as more opportunities are noticed.

When we take opportunities, we freely choose to do so. This freedom of choice is a person's ability to act as she wishes, out
of her perceived options. Freedom of choice is innate, not only to human beings, but to many lower animals - for example, a wolf chooses which deer in a herd to attack. This freedom is unaffected by determinism. And since we have demonstrated that we are responsible creatures without recourse to freedom of choice, we do not have to worry about one standard objection to compatibilism. As noted in chapter II, incompatibilists have argued that being able to act in accordance with one's will, if the will itself is determined, is insufficient for moral responsibility. But incompatibilists argue only that freedom of choice does not guarantee morality, not that we do not have the capacity to choose in accordance with our wills. We have this ability, just as we have the ability of self-reflection that makes us responsible creatures. The two abilities are separate. Freedom of choice is also separate from the practice of taking or assigning responsibility for specific actions, although this latter practice does require freedom of choice. After all, one's will is a part of one's character, and it is the connection between character and action that forms the basis of our assigning responsibility for actions.

People desire both freedom of choice and freedom of opportunity. A person's sense of power over the environment will increase if she sees a number of opportunities to affect it, using a number of her abilities. If she sees no chance to do something that she desires, she suffers a lack of freedom of opportunity. She may still choose among her available options, exercising her
freedom of choice, but will not think this as desirable as perceiving real opportunities to fulfil her currently thwarted wishes.

On the other hand, a person is less likely to be concerned with the lack of freedom of choice, for she can lose her freedom of choice in only a few cases. She might lose all her desires, such as when she is deeply asleep, and thus have no will for actions to be congruent with; or else her actions might somehow be entirely disconnected from her will, as with someone undergoing an epileptic seizure. To lack freedom of choice usually serves as an excuse from responsibility for an action. Since people desire to be responsible beings, they desire freedom of choice. Nevertheless, as has been shown, responsibility consists of more than mere freedom of choice; it is a judgement of one’s character, and of the choices that character leads her to make.

Determinism limits neither the freedom of choice nor that of opportunity. Even if determinism is true, our deliberating process is still causally prior to our actions. We can predict how our actions, and their effects upon the environment, would be different if we had made different choices, and how our choice would be different, had we different characters. Our choices determine our actions. Since we feel that our choices are determined by our character and thus are our own, we feel the same way about our actions. This explains why freedom of choice is important to us, and why determinism is not a threat to it. As for freedom of opportunity, we do not require determinism to be
false in order to perceive the ways in which we can use our performance-abilities. We can be determined to notice opportunities and still retain our freedom of opportunity. We might also be denied opportunities by other people, but determinism does not deny us opportunities in this manner. Thus, our freedom is not restricted by determinism, but by ignorance and coercion.

I have argued that freedom and responsibility are alike in being unaffected by determinism. Yet the relationship between freedom and responsibility is looser than some have supposed. We are responsible beings because we have the general ability to reflect upon our character, which is no more tied to freedom or determinism than is a bird’s general ability to fly. Our ability to perform specific actions, such as scoring a goal, or driving safely, is more closely linked to freedom of opportunity, since it is the latter that allows us to use our abilities. To have opportunities and to see them, we need to be free of coercion and ignorance. We cannot score in hockey if we are blocked by too many opposing players, or if we do not know where the net is. Yet we also need specific performance-abilities, such as the ability to score in hockey. Otherwise, an opportunity to score will be wasted on us.

Although we usually have freedom of choice, and do not have to worry about preserving it, we are expected as human beings to acquire and maintain a certain level of freedom of opportunity. Sometimes, ignorance and coercion do not excuse our mistakes. In
these cases we should have taken steps to become aware of opportunities to avoid restrictive situations, and to develop enough knowledge to see appropriate opportunities. Thus, we are responsible for keeping ourselves fairly free. There are overriding circumstances that can prevent us from maintaining our freedom, but in general we are expected to sustain it at a certain level.

We value our freedom of opportunity. Still, as opposed to the common view that responsibility requires freedom, the converse seems true. Our level of freedom can increase or decrease, but is separate from responsibility. We need only the general ability to modify our character, not freedom of opportunity, in order to consider ourselves fully responsible beings. To be responsible for specific actions we do need freedom of choice, but we can easily meet the latter condition. But we do not need freedom of opportunity in order to be considered responsible people that can legitimately be praised or blamed for their actions. Rather, we have the responsibility to develop characters that can maintain freedom of opportunity; a lack of this freedom will only serve as an excuse up to a point. At bottom it is character, not freedom of opportunity, that matters.

Strawson's account of compatibilism allows for our being responsible agents, and responsible moreover for some actions but not for everything we do. His account is also consistent with our desire, arising from our reactive attitudes, to improve our character, and that one way to improve our character is to
increase our freedom of opportunity. The means by which we fulfil our desire to improve our character is through taking and ascribing responsibility. The better our character is, the more likely that we will have the proper desires and thus exercise our freedom of choice to perform the proper actions. All of this is compatible with determinism. Thus, Strawson's account has led the way to a useful explanation of responsibility and freedom which is unaffected by determinism.
1. For a different view that also ties moral feelings to a definition of responsibility, see R. Jay Wallace, Responsibility and the Moral Sentiments (Cambridge: Harvard University Press, 1994).

2. Taylor, "Responsibility for Self" in Free Will.

3. Dennett, Elbow Room, especially chapter 2.


5. Dennett, Elbow Room, 70.

6. Of course, chapter VIII shows that we sometimes also take into account the results of these strategies. Thus, we blame a drunk driver who hits a child more than one who does not. But we blame both more than we do a sober driver who takes all suitable precautions but accidentally hits a child anyhow.

7. This may explain why Strawson relies upon the infrequency of periods of stress to excuse people of responsibility for them. See "Freedom and Resentment", 79.


9. Relevantly similar actions or situations are those between which there are no detectable relevant differences. Roughly similar actions or situations are those in which there are some detectable relevant differences, but the differences are minor enough that the strategy applicable to the main situation or action should apply to the roughly similar ones as well, and not merely the relevantly similar ones. Thus, if it is wrong to steal bread one time, it is not only wrong to steal bread in general; it is probably wrong to steal anything in general.

10. As Bennett notes, this is where we create an interpersonal relationship with the other, if one does not yet exist. See chapter VII.

11. See Wallace, Responsibility and the Moral Sentiments, 123.

Chapter X: Conclusion

In this thesis I have attempted to defend compatibilism. To this end I have used Strawson's fresh account of compatibilism detailed in "Freedom and Resentment". I had three main objectives. First, to explain Strawson's account. Second, to defend both the necessity and the rationality of our holding reactive attitudes towards ourselves and others, even if determinism is true. Finally, to offer an account of responsibility that gives the reactive attitudes a central role. Hopefully, I have provided in my definition a rough indication of why most people are responsible beings, why such responsibility is desirable, and why people are considered responsible for some events and not for others.¹

There are still questions to explore. I have left open how much leeway we should give agents when we judge the extent of their responsibility for an action, or determine whether or not they are fully responsible agents at all. For instance, I have not said how much ignorance of the consequences of an action we allow a person before we start blaming her. Nor have I attempted to fix the age of majority in order to resolve its variation among different cultures. In these cases and many others there may be an optimum amount of leeway that would both seem intuitively plausible and benefit a society most. I have, however, decided such issues to be beyond the scope of this thesis.

Strawson's account of compatibilism is a significant contribution to the general issue of free will and determinism.
It provides the key to solving the centuries-old problem of how to give an account of moral responsibility in a possibly determined world.
1. There remains the interesting question of how humans came to develop reactive attitudes in the first place. For an enlightening discussion of one possible source of our reactive attitudes, see J.L. Mackie’s "Morality and the Retributive Emotions" in Persons and Values, Joan Mackie and Penelope Mackie, eds., (Oxford: Clarendon Press, 1985).
Bibliography


Hare, Robert D. Without Conscience: The Disturbing World of the Psychopaths Among Us (New York: Pocket Books, 1993).


