View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by South East Academic Libraries System (SEALS)

DETERMINISM AND REACTIVE ATTITUDES: REFLECTIONS ON OUR ALLEGED UNRENOUNCEABLE COMMITMENTS

A thesis submitted in partial fulfilment of the requirements for the degree of

MASTER OF ARTS

of

RHODES UNIVERSITY

by

LINDSAY-ANN KELLAND

March 2009

ABSTRACT

There seems to exist a tension between our metaphysical and phenomenological commitments in the free will debate. On the one hand, I argue that at the metaphysical level we cannot coherently defend the belief that we are morally responsible in the sense that we deserve to be rewarded and punished for our actions, where desert-entailing moral responsibility is the primary understanding of moral responsibility presupposed in the free will debate. I argue that we are responsible for our actions but only in the weaker sense, termed 'attributability' by Gary Watson.

On the other hand, we are allegedly unrenounceably committed at the phenomenological level to conceiving of, and treating, ourselves and one another as morally responsible beings in the desert-entailing sense. P. F. Strawson famously defends this claim in his seminal work, 'Freedom and Resentment'.

In my thesis I will set out this tension by exploring both commitments in turn. I then aim to show that the tension can be dissolved by arguing, contra P. F. Strawson, that our phenomenological commitment is not in fact unrenounceable.

The dissolution of this tension entails, I argue, that we must examine our conception of self and other. We must explore the implications of adopting a position which denies that we are morally responsible beings for our life-hopes, personal feelings, inter-personal relationships and projects. Most importantly, I argue that we must renounce our current retributive condemnatory practices which are based on the unjustified belief that we are morally responsible beings.

TABLE OF CONTENTS

Acknowledgments **iv**

Introduction **1**

Chapter One: Our Metaphysical Commitment **6**

1.1 Libertarianism as an Untenable Position 6

1.2 Hard Determinism 12

1.3 Ted Honderich: Origination and Voluntariness 13

1.4 Compatibilism 14

1.5 Gary Watson: Accountability and Attributability 22

1.6 Voluntariness and Attributability & Origination and Accountability: An Important Relationship 26

1.7 Rejecting Moral Responsibility: Galen Strawson and Thomas Nagel 27

1.8 My Positive Account: A New Position in the Debate 31

Chapter Two: Our Phenomenological Commitment **33**

2.1 The Commonplace Reactive Attitudes and Moral Responsibility 36

2.2 The Objective Attitude 40

2.3 The Commonplace Reactive Attitudes and the general thesis of determinism 41

Chapter Three: Dissolving the Tension **45**

3.1 The ‘Internal’ Argument 46

3.2 The ‘Naturalist’ Argument 48

3.3 The ‘Pragmatic’ Argument 51

3.4 The Fairness of our Moral Practices 53

Chapter Four: The Ramifications **57**

Conclusion **64**

Bibliography **65**

ACKNOWLEDGMENTS

Many thanks to my supervisor Dr Pedro Tabensky, of the Rhodes University Philosophy Department, whose book *Judging and Understanding* inspired me to write this thesis. His energy, support and endless patience have been greatly appreciated. Special thanks to the members of the Rhodes University Philosophy Department and particularly to Sharli-Anne Paphitis, James Vos and Nikolai Viedge without who philosophy would not be what it is. I am indebted to the Andrew Mellon Foundation whose funding has allowed me to pursue my research.

INTRODUCTION

> It is a sort of rape and perversion of logic. But the extravagant pride of man has managed to entangle itself profoundly and frightfully with just this nonsense. The desire for "freedom of the will" in the superlative metaphysical sense, which still holds sway, unfortunately, in the minds of the half-educated; the desire to bear the entire and ultimate responsibility for one's actions oneself, and to absolve God, the world, ancestors, chance, and society involves nothing less than to be precisely this *causa sui* and… to pull oneself up into existence by the hair, out of the swamps of nothingness – Friedrich Nietzsche[1]

> There is no such thing as free will. There is a fundamental sense of the word 'free' in which this is incontrovertibly true… There are plenty of senses of the word 'free' in which it is false. But the sense in which it is true seems to be the one that matters most to most people – at least in so far as questions of morality are concerned. Or rather, it seems to be the one that most people think matters to them – rightly or wrongly. – Galen Strawson[2]

Our experience of ourselves as morally responsible beings cannot properly be accounted for by our best theories of free will. It seems that there is a very real tension between our conceptions of ourselves in this regard and what we can properly show to be the case. In this thesis I will explore this supposed tension and show, among other things, that the tension is expressive of two contradictory commitments. On the one hand, I will defend the somewhat controversial view that the best position in the free will debate combines the insights of compatibilism with those of hard determinism. This view denies that we are morally responsible agents but still defends the claim that we have free will (even if not in the superlative metaphysical sense that Nietzsche rejects).[3] On the other hand, one thing that is regularly highlighted in the free will debate, particularly but not exclusively by hard determinists, is the recalcitrance of our commonplace reactive attitudes – resentment, indignation, gratitude and so on.[4] These attitudes are constitutive of our deeply held belief that we are indeed morally responsible agents. It seems that the free will we experience ourselves as having implies the belief that we are morally responsible and yet reason forces us to conclude that we cannot, in principle, be morally responsible for what we do. So, given this

[1] Nietzsche, Beyond Good and Evil, §21.
[2] Strawson, G. *Freedom and Belief*, v.
[3] I am aware that moral responsibility is standardly seen as central and indeed necessary for free will. I will be arguing against this mainstream view.
[4] This view is classically expressed by Peter Strawson in his seminal work 'Freedom and Resentment', found inWatson, G (ed.) *Free Will: Second Edition*, Oxford University Press, pp. 72-93.

tension, I will examine the concept of free will in two distinct ways: metaphysically and phenomenologically. This dual-aspect analysis of free will aims to show how deeply problematic the tension between our metaphysical and phenomenological commitments is. Once this has been shown I will put forward and defend a resolution of this tension.

Chapter Breakdown

In Chapter One I explore our metaphysical commitment in the free will debate. There are three traditional positions in the debate: libertarianism, compatibilism and hard determinism. Libertarians defend the claim that we are free and morally responsible agents. The freedom they endorse involves self-determination – to be the first and primary cause of one's actions – and this freedom, they argue, makes us morally responsible in the sense that we deserve to be praised or blamed and rewarded or punished for our actions. Hard determinists, on the other hand, argue that free will and moral responsibility are impossible because self-determination cannot be made coherent under either determinism – the thesis that the state of the world at any given time is fixed by prior states or by the laws of nature – or indeterminism – the thesis that at least some states of the world are not fixed in this way. Finally, compatibilists argue that the possible truth of determinism has no bearing on our free will or moral responsibility and put forward accounts of free will that are not grounded in self-determination but rather in structural features of the will. All three positions assume that desert-entailing moral responsibility is central and indeed necessary for free will.

First, I rehearse three well-known objections to libertarianism. I assume, uncontroversially, that these objections are correct and that they show libertarianism to be an untenable position in the debate. Second, I examine the strengths and weaknesses of compatibilism and hard determinism and argue that neither account by itself is able to capture who we are as agents. Given this, I argue that we should be committed to a new position which draws important insights from both the compatibilist and hard determinist camps. I argue that while we do have free will we are not morally responsible in the desert-entailing sense. Although I am aware that desert-entailing moral responsibility is seen as central and indeed necessary for free will, as mentioned above, I will be arguing against this mainstream view. I argue

instead that the only type of responsibility which we can plausibly defend is attributability, a distinct and weaker type of responsibility described by Gary Watson, arguably one of the most influential compatibilists currently working in the free will debate. This conception of responsibility is *prima facie* compatible both with compatibilism and hard determinism and unlike the desert-entailing variety also happens to be invulnerable to sceptical attacks from hard determinists. Moreover, it is the only type of responsibility we actually possess given the types of agents we are. In defending this view I part company with compatibilists and hard determinists alike. Once it is acknowledged that what divides the compatibilist and hard determinist camps hinges on the idea of desert and that desert should be left by the wayside, it seems that it is most plausible for us to be committed to a position which combines the insights of both camps. Our best accounts of agency and freedom, then, do not provide us with the desert-entailing moral responsibility we typically experience ourselves as having.

In Chapter Two I examine Peter Strawson's seminal paper, 'Freedom and Resentment'[5]. This paper points to the basis of the second commitment in question: our phenomenological commitment to our so-called commonplace reactive attitudes. According to Strawson, the reactive attitudes are constitutive of who we are as agents, of our intra- and inter-personal relationships and of the moral community itself. Given this, he argues that we have an unrenounceable commitment to the reactive attitudes since to renounce them would cause the collapse of the moral community and end all of our personal relationships. A metaphysical analysis (like that presented in Chapter One) is, for Strawson, completely irrelevant to our justification of these concepts; our unrenounceable commitment to the reactive attitudes justifies our beliefs in freedom and moral responsibility because it is only in terms of the commonplace reactive attitudes that we can explain how it is that we are free and morally responsible. For Strawson, that is, it is only appropriate to respond to another with the reactive attitudes if we see the other as morally responsible, and, he argues, the appropriateness of our response means that the other is in fact morally responsible. According to Strawson, the very concepts of freedom and moral responsibility only make sense in light of our experience of ourselves, one another, and the world. I argue

[5] Watson, G (ed.) *Free Will: Second Edition*, pp. 72-93.

that if Strawson is correct then we seem to have an irresolvable tension between these two commitments: our metaphysical commitment which denies that we are morally responsible and our phenomenological commitment which justifies our belief in moral responsibility.

In Chapter Three I argue that the tension is in fact resolvable and suggest a way of dissolving it. I explore Galen Strawson's response to his father in 'On "Freedom and Resentment"'.[6] Here G. Strawson, a hard determinist, argues that despite the fact that we seem to have an unrenounceable commitment to the reactive attitudes this commitment does not provide the *right sort* of justification for our beliefs in freedom and desert-entailing moral responsibility and, therefore, does not show the hard determinist's concern for *metaphysical* justification of these concepts to be misplaced. For G. Strawson, without showing that we actually are free and morally responsible at the metaphysical level we cannot justify our believing just this. However, by acknowledging our commitment to the reactive attitudes G. Strawson buys into the idea of an irresolvable tension between our two conflicting commitments and describes this tension as "a very real conflict of commitment".[7] He claims that:

> While we have a deep and perhaps inderacinable commitment to the reactive attitudes and practices, it is also in our nature to take determinism to pose a serious problem for our notions of responsibility and freedom… our commitments are complex and conflict.[8]

Through an examination of R. Jay Wallace's *Responsibility and the Moral Sentiments*[9] I argue, contra G. Strawson, that the tension can be dissolved. My argument is developed in two distinct ways. First, I argue that some of the reactive attitudes – the retributive reactive attitudes – are not in fact commonplace and can be renounced without causing the collapse of the moral community or meaning the end of all intra- and inter-personal relationships. I do this by showing that P. F. Strawson's three arguments defending our commitment to the commonplace reactive attitudes against the threat of determinism are flawed. Second, I argue we ought examine the fairness of our retributive practices which are founded on the unjustified belief that we are morally responsible beings. It seems, I will argue, that our commitment to truth and reason ought to trump our practically basic phenomenological commitment to

[6] Fischer, J. M. & Ravizza, M. (eds.) Perspectives on Moral Responsibility, pp. 67-100.
[7] Strawson, G. 'On "Freedom and Resentment"', p. 72.
[8] Ibid, p. 71.
[9] Wallace, R. Jay. *Responsibility and the Moral Sentiments.* 1998.

experiencing and treating ourselves and one another as morally responsible agents. Our commonplace reactive attitudes, that is, express beliefs that we hold, where to believe *x* means that we take *x* to be true. To elucidate, we believe that it is wrong to commit murder. This belief expresses what we take to be a moral obligation, which if breached typically causes members of most societies to respond retributively with the reactive attitude of resentment. In this case our resentment towards the perpetrator expresses our belief that it is wrong to commit murder as well as the further belief that the perpetrator deserves to be punished for his action – is morally responsible for committing the murder in the desert-entailing sense. I have used the example of resentment because this attitude, in expressing the belief that we are morally responsible in the desert-entailing sense is incommensurable with what we are able to justify at the metaphysical level. Reason forces us to choose between our conception of ourselves as morally responsible agents and our best accounts of agency. Faced with this choice, and the implication that our moral practices are unfair, I argue that our metaphysical commitment which denies that we are morally responsible agents must trump our conception of ourselves as precisely this. Taking ourselves to be morally responsible agents is entirely different from actually being morally responsible agents and in light of this knowledge we ought to examine, reflect on and criticise our beliefs about ourselves as well as the practices we found on these beliefs.

In the final chapter I briefly set out two other non-standard positions in the free will debate which, like my own, combine insights from both the compatibilist and hard determinist camps. These positions are put forward by Saul Smilansky and Derk Pereboom. While the metaphysical commitment I advocate in this thesis wholly breaks with tradition there is, given these positions, a movement in this direction. Finally, through my discussion of Pereboom's paper I provide an indication of the actual consequences of adopting a position which denies that we are morally responsible in the desert-entailing sense. Pereboom explores the consequences of adopting such a position on our life-hopes, personal feelings, view of ourselves and view of others. Finally, I argue, contra P. F. Strawson, that our retributive condemnatory practices ought to be renounced given that they are founded on our unjustified belief in desert-entailing moral responsibility.

CHAPTER 1: OUR METAPHYSICAL COMMITMENT

> Our sense of ourselves as free is perplexing not only because of hard questions about necessity and contingency, actions and events, reasons and causes. The perplexity is ethical as well as metaphysical. Our view of the nature of free will is bound up with our view of why free will matters." – Gary Watson.[10]

1.1 Libertarianism as an Untenable Position

Freedom of the will and moral responsibility seem to matter greatly to us and seem to play a very important role in our lives. We praise those who do great things and imprison transgressors to a large extent, but not exclusively, because we consider them to be morally responsible for their wrongdoings and, relatedly, because we think of them, prototypically, as able freely to choose.[11] We think of them as morally responsible insofar as we think of them as praise or blameworthy for their actions in the sense of rightly deserving reward or punishment. The concepts of praise and blame, which are standardly taken to be inseparably tied to the idea of moral responsibility, are clearly expressed in certain widespread religions where believers are promised either eternal bliss in heaven or eternal damnation in hell based on their actions here on earth. Although the concepts of praise and blame are not exclusively religious concepts, the heaven-and-hell metaphor exemplifies the retributive core of our reward and punishment practices. As Galen Strawson puts it:

> True moral responsibility is responsibility of such a kind that, if we have it, then it *makes sense*, at least, to suppose that it could be just to punish some of us with (eternal) torment in hell and reward others with (eternal) bliss in heaven… The story of heaven and hell is useful simply because it illustrates, in a peculiarly vivid way, the *kind* of absolute or ultimate accountability or responsibility that many have supposed themselves to have, and that many do still suppose themselves to have.[12]

Although there are, arguably, various conceptions of responsibility, this is the definition that we should be working with when we speak about true desert-entailing *moral* responsibility – the idea of being praise or blameworthy for one's actions in the sense of deserving reward or punishment – as this is the primary understanding of

---

[10] Watson, G. *Free Will: Second Edition*, p. 1.

[11] I add the qualifiers 'but not exclusively' and 'prototypically' here because when we imprison a person who has committed murder but who also suffers from, say, paranoid schizophrenia we do not think of her as truly morally responsible for her action because we do not see her as being able to make choices in the way that we typically are.

[12] Strawson, G. 'The Impossibility of Moral Responsibility', p. 515.

moral responsibility presupposed in the free will debate insofar as the different position in the debate, primarily libertarianism, hard determinism and compatibilism, are carved out in relation to this conception. While I deny that we are morally responsible in this desert-entailing sense, I will defend a weaker type of responsibility termed attributability by Gary Watson and claim that attributability is the only type of responsibility we can coherently defend and the only type of responsibility which we need in order to have wills that are free. In my discussion of libertarianism however, when I speak about moral responsibility I am speaking about *desert-entailing* moral responsibility.

Immanuel Kant argues that:

> Man *himself* must make or have made himself into whatever, in a moral sense, whether good or evil, he is to become. Either condition must be an effect of his own free choice; for otherwise he could not be held responsible for it and could therefore be *morally* neither good nor evil.[13]

Here Kant is arguing that if we are properly to be held morally responsible for our actions then we must be responsible for who we are and this can be the case if *and only if* we have made ourselves the way we are – if we are self-determining. This conception of freedom is assumed by all incompatibilists, libertarians and hard determinists alike. Unlike hard determinists, however, libertarians argue that we are free in the sense endorsed by Kant in the quote cited above and are therefore they argue morally responsible in the desert-entailing fashion. Libertarians, that is, argue that we are self-determining agents. To be self-determining, or *causa sui*, is to satisfy Kant's criteria for moral responsibility precisely because being *causa sui* means that one is the cause of oneself and the first or primary cause of one's actions.

In what directly follows let us assume, for the sake of argument, that incompatibilists are correct to argue that we can only properly be held morally responsible if we are self-determining *causa sui* or, to use another familiar expression coined by Ted Honderich, the originators of our actions. For the sake of argument let us also assume in what directly follows that libertarians are correct to argue that we are free in the way outlined by Kant in the quote above and are therefore morally responsible in the desert-entailing fashion.

In what follows I hope to show that it is impossible for us to be self-determining. I will do this by arguing that the concept of self-determination is effectively

[13] Kant, I. *Religion within the Limits of Reason Alone*, p. 40.

challenged on three fronts: first by determinism and second by indeterminism – where these two theses are opposed to one another, that is mutually exclusive, and exhaustive. Third, the concept of self-determination is effectively challenged by Galen Strawson who attacks self-determination in 'The Bounds of Freedom'[14], 'The Impossibility of Moral Responsibility'[15] and elsewhere. In what follows I will briefly explore all three challenges. I aim to show that it is impossible for us to be self-determining so that I can show libertarianism to be an untenable position in the debate and focus my attention on what are currently the two widely debated positions in the free will debate: compatibilism and hard determinism. The challenges set out below, then, are currently standardly accepted as sufficient to rule out the possibility of libertarian free will.[16]

Firstly, then, the idea of self-determination is challenged by the general thesis of determinism as defined above. If determinism is true then it seems as though we cannot be said to be the originators of our actions at all. Rather, it seems as though we are merely passive participants in the causal schema. If this is the case, however, then it is hard to see how we possess free will or can be seen as morally responsible for our actions. In fact, when we first encounter the thesis of determinism, and imagine it to be true, it seems hard to see how we can be seen as agents *simpliciter*. Determinism, that is, is straightforwardly incompatible with self-determination. The concern that determinism straightforwardly undermines agency and free will is addressed by compatibilism. Compatibilism is the view that free will can be explained without appealing to the concept of self-determination, which opens the way to being able to provide an account of free will that is allegedly compatible with determinism. I will discuss this view extensively later in this chapter. At this stage in my argument my main concern, as mentioned above, is to show why libertarianism is untenable, since showing this allows to focus on compatibilism and hard determinism and the insights and shortfalls of both views.

---

[14] Kane, R (ed.) *The Oxford Handbook of Free Will*, pp. 441-460.
[15] Feinberg, J & Shafer-Landau (eds.), R, *Reason and Responsibility*, pp. 513-522.
[16] I am aware that there are some recent relatively sophisticated defences of libertarianism, which I have no space to explore here (e.g. Kane, Wiggins) and which I think ultimately fail. My aim here is relatively modest. I aim solely to establish a *prima facie* case against libertarianism. If the arguments presented are not strong enough then my conclusions should be thought of as provisional. But, that said, and despite the recent valiant attempts to resuscitate libertarianism alluded to above, this position in the free will debate has largely been discredited.

The second challenge mentioned above is the challenge of indeterminism, the converse of determinism. Indeterminism is the thesis that some events are uncaused by events or the state of the world prior to them. If we reject determinism in favour of indeterminism in an attempt to find a place for agency and free will, we still find ourselves unable to ground a coherent metaphysical picture of self-determination. Indeterminism, that is, seems equally incompatible with the basic notion of *causa sui*. To see why this is so we need to explore a typical objection levelled at libertarians, namely the objection that their accounts of free will entail that our choices are ultimately arbitrary. The objection can be better understood if we examine Donald Davidson's claim, famously defended in his paper 'Actions, Reasons, and Causes', that reasons are causes. The view expressed by Davidson is standardly accepted in the free will debate, so here I will, relatively uncontroversially, assume that it is right. To say that reasons are causes is to say that the reasons we have for acting the way we do cause us to act as we do. This claim holds clout because intuitively it seems as though we must be able to provide explanations for the way we act if we hope to show that our actions are indeed ours and it seems plausible that we can only truly explain how it is that we are moved by mental states such as beliefs and desires (reasons) if we attribute causal powers to mental states. As Davidson writes:

> What is the relation between a reason and an action when the reason explains the action by giving the agent's reason for doing what he did? We may call such explanations *rationalizations*, and say that the reason *rationalizes* the action… I… defend the ancient – and commonsense – position that rationalization is a species of causal explanation.[17]

To elucidate, if I believe that adding milk and sugar to tea makes it taste better then my doing so is the direct result of this belief; my belief that adding milk and sugar to tea makes it taste better *causes* me to do just this when I make tea.

Now recall that indeterminism is the thesis that some events are uncaused by events or the state of the world prior to them. According to the standard argument, we would be unable properly to explain why or how our actions flow from our reasons, or are indeed ours in the sense outlined above, in an indeterministic framework. For this reason it is typically objected that indeterminism entails that our choices are arbitrary and our actions ultimately random. If we assume that libertarians are correct, as we have for the sake of argument, then indeterminism is incompatible with self-determination and moral responsibility because it entails that our actions occur

[17] Davidson, D, 'Actions, Reasons, and Causes', p. 675.

randomly, and if our actions occur randomly then they cannot be said to originate within us – we cannot be said to be the first or primary cause of them.

We can clarify the problem of arbitrariness, as well as the related idea that reasons are causes, if we examine a thought experiment provided by Pedro Tabensky.[18] Imagine a case of identical twins – with exactly the same past, genetic makeup, values, beliefs, preferences, desires and histories – who are presented with a choice between two distinct alternatives, A and B, at a given point in time. Given our pre-reflective, unconsidered beliefs about freedom, namely that we are self-determining agents, we want to be able to say that it is possible for to choose A and to choose B. However, given that *everything* about the twins is exactly the same we need to question what would prompt their different choices at this point. What accounts for choosing A and choosing B? Since both twins, by stipulation, have exactly the same mental states, and it seems as though our mental states cause us to act as we do, there seems to be nothing that could explain the difference in choice. The different choices are wholly unexplainable and mysterious. That is, if we take seriously the fact that in every respect the twins are identical then it should be clear that there exists no plausible explanation for the difference in choice. But this entails that the choices made at this point would have to be fundamentally random or arbitrary and therefore not expressive of human agency.

Indeed, looking at the problem from a slightly different angle, speaking of identical twins, of the extreme sort that have just been described, amounts, for our purposes, to much the same as speaking of a single person faced with a choice between A or B. The thought experiment shows that self-determination or origination – that in any *typical* situation of choice at least two options are genuinely available – is implausible. If indeed the twins were able to choose differently in the sense that the choice originates in them, if a single person is able to choose either A or B in this way, then their choices would in an important sense not be theirs at all, which is to say that their choices would not be choices proper. And if their choices are not choices proper, then it is hard to see how they could be thought of as flowing from a will that is free at all. If we are free, we need to make sense of the basic idea that our choices

[18] Pedro Tabensky, 2006. This case was discussed in conversation and does not appear in print.

and actions flow from relevant states of mind and it seems as though the only way to do this is to think of reasons as causes.

Libertarians have responded to the problem of arbitrariness by putting forward notions such as agent-causation (Chisholm) and self-forming actions (Kane)[19] but these attempts have widely and effectively been criticised for falling prey once again to the very same problem of arbitrariness and mystery.

It could still be argued, contra what I have said above, that while our actions may be caused by our relevant mental states, this is not to say that they are *determined* by such states. In response, it could in actual fact be that the universe is indeterministic, but it seems as though what allows us to pick out the agency of agents is a good causal story, and so it is the case that we could assume, for our purposes, the thesis of determinism, for where agency is concerned it would not matter whether the universe is deterministic or not. What does matter, for our purposes, is that it seems as though the only means we have of explaining human actions is in terms of causal accounts and, given this, it seems as though what makes us free agents, if anything does, is that our actions *qua* actions can be understood as caused by our mental states. But this mode of explaining human action does not, contrary to the compatibilist position which shall be discussed further on, allow us properly to account for the deeply held belief that we are morally responsible in the desert entailing fashion.

Finally, Galen Strawson in both 'The Bounds of Freedom' and 'The Impossibility of Moral Responsibility' reduces the concept of self-determination to absurdity. G. Strawson, like Kant, claims that to be self-determining "one must have consciously and explicitly chosen to be the way one is, mentally speaking, in certain respects, *and* one must have succeeded in bringing it about that one is that way".[20] But, he argues, in order to choose the way I am mentally speaking I must already have certain beliefs and preferences *x* in light of which I choose how I want to be. But if I am to be truly self-determining then I must have chosen *x* in light of another set of beliefs and preferences *y*. However, for the same reason I must have chosen *y* in light of yet another set of beliefs and preferences *z* and so on. The sets of beliefs and

[19] Chisholm, R. M. 'Human Freedom and the Self' and Kane, R. 'Free Will: Ancient Dispute, New Themes' both in Feinberg, J & Shafer-Landau, R (eds.) *Reason and Responsibility*, pp. 492-498 and pp. 499-512.
[20] Strawson, G. 'The Impossibility of Moral Responsibility', p. 513 (my emphasis).

preferences which I must possess in order to have consciously and explicitly chosen the way I am can continue in this way to infinity. This is known as an infinite regress problem. Strawson's challenge to self-determination, then, is that in order to be *causa sui* we must consciously and explicitly choose the way we are but this is impossible since choosing the way I am in this way implicitly involves an infinite regress. As Galen Strawson puts it:

> Here we are setting out on a regress that we cannot stop. True self-determination is impossible because it requires the actual completion of an infinite series of choices of principles of choice… Nothing can be *causa sui* in the required way.[21]

In short, determinism challenges the idea of self-determination since the very thesis of determinism is straightforwardly incompatible with the concept of self-determination. Indeterminism, the converse of determinism, also challenges the idea of self-determination because it seems to entail that our choices are arbitrary and our actions random, an entailment that is incompatible with self-determination because it rules out the idea that actions flow from agents. However, since determinism and indeterminism are exhaustive options, and both are incompatible with self-determination, it seems impossible for us to be self-determining.[22] This impossibility is reinforced by Galen Strawson's *reductio* of self-determination which still remains to be refuted. For the reasons provided above, libertarianism – because it depends on self-determination – is untenable and can with relative safety be moved aside, leaving only compatibilism and hard determinism (or, as I will argue, a hybrid of the two) as genuinely live options in the free will debate.

1.2 Hard Determinism

Hard determinists, like libertarians, think of free will as grounded in self-determination, or being *causa sui.* Also like the libertarians, hard determinists see free will and moral responsibility as incompatible with determinism. Hard determinists, however, argue that it is impossible for us to be self-determining agents. Free will and moral responsibility, then, are for the hard determinist nothing more than illusion.

---

[21] Ibid, p. 514. Here 'principles of choice' refers simply to the set of preferences one has to have in light of which we choose how we want to be.

[22] Later I will defend the incompatibilist's claim that moral responsibility is inseparably tied to self-determination; that we must be self-determining if we are *properly* to be held morally responsible for our actions. However, given the impossibility of self-determination, I will argue that moral responsibility too is impossible. In this regard I am at one with hard determinists.

According to the hard deterministic, whenever we are presented with a choice between, say, A and B, there is, given the antecedent conditions at that precise moment in time, only ever one genuine option available to us. We are never able to do other than what we did – except in the merely counterfactual way where I could have done otherwise if my antecedent conditions at the time were relevantly different. In s Ted Honderich's words, "An open future, a future we can make for ourselves, *is* one of which determinism isn't true".[23] Similarly Paul Holbach, a hard determinist writing in the century, writes:

> In whatever manner man is considered, he is connected to universal nature, and submitted to the necessary and immutable laws that she imposes on all the beings she contains… Man's life is a line that nature commands him to describe upon the surface of the earth, without his ever being able to swerve from it, even for an instant… Nevertheless, in spite of the shackles by which he is bound, it is pretended he is a free agent, or that independent of the causes by which he is moved, he determines his own will, and regulates his own condition.[24]

1.3 Ted Honderich: Origination and Voluntariness

Honderich draws a distinction, crucial for our purposes, between origination and voluntariness. For Honderich, if I am the originator of my actions I am self-determining or *causa sui*. In contrast to this, voluntary actions, which are compatible with both compatibilism and hard determinism, are actions that flow naturally and directly from our desires, beliefs, intentions and so on. The principal difference between compatibilism and hard determinism is that compatibilists see voluntariness as sufficient for free will and desert-entailing moral responsibility while hard determinists argue that we can only be said to be free or morally responsible if we are self-determining – if we originate our actions and this, they claim, is impossible. As Honderich writes:

> In the Philosophy of Mind itself, we find only philosophers who assume or explain that human choices and actions are effects of causal sequences or chains of the sort that are taken in the literature on determinism and freedom to raise the further question of our freedom. When philosophers are concerned with consciousness and mental activity… they have nothing to say of origination… Origination's absence from the Philosophy of Mind can indeed be taken to suggest that there is no *tempting* conception of origination in existence. Otherwise it would certainly have been made use of in general explanations of behaviour.[25]

Hard determinists, I will argue below, are correct to claim that we are not morally responsible in the desert-entailing sense. I will also argue that where their

[23] Honderich, T. 'A Defense of Hard Determinism', p. 473.
[24] Holbach, P. 'The Illusion of Free Will', p. 462.
[25] Honderich, T. 'Determinism as true, Both Compatibilism and Incompatibilism as false, and the real problem', pp. 467-468.

arguments fail is in assuming, contrary to compatibilists, that we can only be said to have free will if we are self-determining.

1.4 Compatibilism

Compatibilists, as opposed to incompatibilists, argue that free will and moral responsibility are compatible with determinism. For the compatibilist, freedom of the will is not to be located in self-determination but rather in structural features of the will which opens the way to providing accounts of free will that are allegedly compatible with determinism. What matters for the purposes of agency, according to the compatibilist, is that our actions are caused by our mental states – that we are able to provide causal explanations for our actions. Recall Davidson's claim that reasons are causes and that what makes us agents at all is that our actions *qua* actions can be understood as caused by our mental states. The compatibilist endorses something like Davidson's position, namely that reasons (and other mental states) must be the cause of action if action is to be understood as action proper. For the compatibilist, then, since reasons are causes, free will should not be contrasted with causation but rather with constraint. A. J. Ayer writes:

> But now we must ask how it is that I come to make my choice. Either it is an accident that I choose to act as I do or it is not. If it is an accident, then it is merely a matter of chance that I did not choose otherwise; and if it is merely a matter of chance that I did not choose otherwise, it is surely irrational to hold me morally responsible for choosing as I did. But if it is not an accident that I choose to do one thing rather than another, then presumably there is some causal explanation of my choice: and in that case we are led back to determinism... It seems that if we are to retain this idea of moral responsibility, we must either show that men can be held responsible for actions which they do not do freely, or else find some way of reconciling determinism with the freedom of the will... it is not, I think, causality that freedom is to be contrasted with, but constraint.[26]

If one is relevantly constrained one does not have sufficient control over what motivates one to action and so, according to the compatibilist, cannot be free. For the compatibilist, then, we have free will when we have control over what moves us to act. The issue of control is explored in different ways by different compatibilists (all of whose accounts can, for the most part, be seen as complementary). I will spend a lot of time exploring two compatibilist accounts of free will below, namely the

[26] Ayer. A. J. 'Freedom and Necessity', pp. 483-484.

accounts of Harry Frankfurt and Gary Watson, arguably the two most influential contemporary compatibilists working on the free will debate.[27]

The control we have over what moves us to action is spelt out by Frankfurt as our ability to endorse, or identify with, those desires which we most want to be our will and by Watson as our capacity to make our valuational and motivational systems correspond so that we are motivated to act only by those courses of action which we deem valuable. Both Frankfurt and Watson, whose accounts will be explored in detail below, see critical reflection as necessary for free will and imagine, what Susan Wolf terms, a 'deeper self' critically reflecting and identifying with certain desires and values. In the following quote Susan Wolf succinctly explores the idea of this deeper self. She says:

> Frankfurt's and Watson's accounts may be understood as alternate developments of the intuition that in order to be responsible for one's actions, one must be responsible for the self that performs these actions… [they] share the idea that responsible agency involves something more than intentional agency… if we are responsible agents, it is not just because our actions are within the control of our wills, but because, in addition, our wills are not just psychological states *in* us, but expressions of characters that come *from* us, or that at any rate are acknowledged and affirmed *by* us. For Frankfurt, this means that our wills must be ruled by our second-order desires; for Watson, that our wills must be governable by our system of values… Because, at one level, the differences [between] Frankfurt [and] Watson… may be understood as differences in the analysis or interpretation of what it is for an action to be under the control of this deeper self, we may speak of their separate positions as variations of one basic view about responsibility.[28]

Compatibilists, like hard determinists and libertarians, assume that one is only morally responsible if one is morally responsible in the desert-entailing sense, and that one can only have free will if one is morally responsible. Compatibilists believe that by providing us with what are arguably our best accounts of free will they are providing us with justification for our belief in desert-entailing moral responsibility. Remember that I will argue against this mainstream view. Although I believe that free will is indeed located in structural features of the will I claim that this type of free will does not justify our belief in desert-entailing moral responsibility. Contra the mainstream, I argue that we are only responsible in the weaker sense termed "attributability" by Gary Watson and argue that this sense of responsibility is sufficient to ground a rather natural understanding of free will and agency. Attributability, I argue, is the only type

[27] I spend more time discussing compatibilism than I did hard determinism because I will argue that compatibilist accounts of free will and agency are the best and most sophisticated accounts currently put forward in the free will debate. It is therefore important to have a more detailed understanding of these accounts.

[28] Wolf, S. 'Sanity and the metaphysics of responsibility', pp. 49-50.

of responsibility which we can properly defend at the metaphysical level and the only type of responsibility which we actually possess. The compatibilist sense of free will is exemplified in the accounts put forward by Frankfurt and Watson; accounts which locate free will in the structural features of the will. I will briefly explore both accounts below to give the reader a better idea of the type of free will I endorse.

Harry Frankfurt provides a hierarchical account of persons as necessarily complex motivational creatures in his seminal work 'Freedom of the Will and the Concept of a Person'.[29] What separates us from other animals, he argues, is critical reflection – we are the types of creatures who are able to care about what motivates us to action. To elucidate, Frankfurt provides an account of first and second-order desires, second-order volitions and the will. A volition differs from a desire in that it is a critically-reflected-wish for a certain first-order desire to be one's will, where the will should be understood as that which motivates one to action – the effective desire that would push one into action if one was able to act. So I could have a first order desire to write my thesis – I want to write my thesis. If I want this desire to be effective in action, to be my will, then I have formed a second-order volition with regards to the desire, where this process necessarily entails that I have critically reflected on the desire. There is a subtle but important difference between a second-order volition and a second order desire. The difference becomes quite clear if we examine Frankfurt's example of the physician who wants to understand his drug-addicted patients' desire to take drugs. The physician, in this case, has a second-order desire to take drugs – he *wants to* desire to take drugs, but this desire is not a second-order volition – he does not want this desire to motivate him to action, or put another way he does not endorse this desire as that which he wants to be his will. Although he wants to desire to take drugs (at the second-order) he does not want to take drugs (at the first-order).[30] Although forming a second-order desire does not necessarily entail that I have critically reflected as is the case when a second-order volition is formed, Frankfurt is only interested in those second-order desires which have involved the process of critical reflection because critical reflection, he claims, is necessary for autonomy and freedom of the will. He says: "No animal other than man… appears to

[29] *The Journal of Philosophy*, Vol. 68, No. 1 (Jan 14, 1971), pp. 5-20.
[30] Frankfurt, H. 'Freedom of the Will and the Concept of a Person', p. 9.

have the capacity for reflective self-evaluation that is manifested in the formation of second-order desires".[31]

For Frankfurt we are capable of having freedom of the will because we have this complex motivational structure, are capable of critical reflection, and have a certain degree of control over which of our desires we want to be our will – we are, he claims, able to have the will we want. He says:

> The statement that a person enjoys freedom of the will means… that he is free to will what he wants to will, or to have the will he wants… It is in securing the conformity of his will to his second-order volitions, then, that a person exercises freedom of the will.[32]

This complex motivational structure is, according to Frankfurt, a necessary condition for personhood. To be a person one must, according to Frankfurt, care about what moves one to action and to care about this entails, as mentioned above, that one has critically reflected on one's first-order desires and through the process of identification formed second-order volitions regarding what one wants one's will to be.[33] Identification, then, takes place at the volitional level.

According to Frankfurt, it is possible for a person to experience volitional conflict. In his papers 'The Faintest Passion' and 'Identification and Wholeheartedness',[34] Frankfurt introduces the concepts of ambivalence and wholeheartedness to expand upon his discussion of volitions and volitional-conflict. He claims that a person is wholehearted if, in the case of a volitional conflict, she has truly resolved which second-order volition she most wants to be her will. Wholeheartedness, then, concerns the organisation of a person's will. As Frankfurt puts it:

> It is these acts of ordering and rejection – integration and separation – that create a self out of the raw materials of inner life… Deciding plays an important role in the formation and maintenance of the self.[35]

Correspondingly, a person is ambivalent if she does not know how to resolve a volitional conflict; she both desires something and its opposite and cannot identify

---

[31] Frankfurt, H. 'Freedom of the Will and the Concept of a Person', p. 7.
[32] Ibid, p. 15.
[33] Frankfurt contrasts the person with the wanton who does not form second-order volitions concerning any of her first-order desires because she simply does not, or is unable to, care about which of her desires ultimately move her to action.
[34] Frankfurt, H. *Necessity, Volition and Love*, 95-107 and *The Importance of What we Care About*, pp. 159-176.
[35] Frankfurt, H. 'Identification and Wholeheartedness', pp. 170-172.

what she wants her will to be. Importantly, Frankfurt argues that ambivalence cannot be overcome voluntarily. He claims that:

> a person's will is real only if its character is not absolutely up to him… we do not control, by our voluntary command, the spirits within our own vasty deeps. We cannot have, simply for the asking, whatever will we want… we cannot be authors of ourselves... we can only be what nature and life make us, and that is not so readily up to us. [36]

One's will, then, is either wholehearted or ambivalent and we have no direct voluntary control over which it turns out to be. However, Frankfurt argues that our lack of control over the *nature* of our will itself does not entail that we do not have control over what we want our will to be. We still control which of our desires we endorse or identify with, as described in 'Freedom of the Will and the Concept of a Person', and this, according to Frankfurt, is enough to give us free will.

At this point the incompatibilist might object, arguing that unless we can control whether our wills are in fact wholehearted or ambivalent we cannot be said to be genuinely free. However, according to Frankfurt, the incompatibilist's assumption that only self-determination can ground free will is not only unrealistic but moreover impossible. He says:

> … what is freedom of the will? A natural and useful way of understanding it is that a person's will is free to the extent that he has whatever will he wants. Now if this means that his will is free only if it is under his *entirely* unmediated voluntaristic control, then a free will can have no genuine reality; *for reality entails resistance to such control*… The dilemma can be avoided if we construe the freedom of someone's will as requiring, not that he originate or control what he wills, but that he be wholehearted in it. If there is no division within a person's will, it follows that the will he has is the will he wants. His wholeheartedness means exactly that there is in him no endogenous desire to be volitionally different than he is. Although he may be unable to create in himself a will other than the one he has, his will is free at least in the sense that he himself does not oppose or impede it.[37]

Frankfurt wears his anti-libertarianism on his sleeve when he claims that it is impossible for a person to have complete control over the nature of her will; reality, for Frankfurt, rules out the possibility of our being self-determining. For Frankfurt (as for all compatibilists) we do not need to appeal to self-determination in order to show that we have wills that are free; free will is already grounded in the limited amount of control which we do in fact have, namely that which extends only to which of our desires we choose to endorse or identify with. According to Frankfurt, we can, for the most part, control the formation of our second-order volitions, and when our actions flow naturally from the volitions we have formed through the critical reflection we

[36] Frankfurt, H. 'The Faintest Passion', p. 101.
[37] Frankfurt, H. 'The Faintest Passion', pp. 101-102 (my emphasis).

can be said to be acting of our own free will and are thereby, Frankfurt thinks, morally responsible for the way we act.

Frankfurt's famous example of the unwilling addict elucidates the type of control we must have in order to be free and morally responsible by exemplifying a case of an unfree person. The unwilling addict tries desperately to avail himself of his addiction to drugs. He has conflicting first-order desires – one to take drugs and the other to refrain from doing so. Because he cares about what his will is and has formed a second-order volition to the extent that he does not want to take drugs he can indeed be considered a person. However, because he cannot overcome his addiction, he cannot be seen as exercising his freedom because he does not have the level of control in this instance necessary to be considered free. Recall that for Frankfurt, freedom of the will "is exercised in securing the conformity of [one's] will to [one's] second order volitions".[38] Because the unwilling addict cannot make himself act in accordance with his own wishes he lacks freedom of the will *in this instance* and cannot therefore be considered morally responsible for taking the drug. The idea of the unwilling addict hangs nicely with our intuitions. When we come across a *genuinely* unwilling addict we do seem to feel pity for her and indeed tend to excuse behaviour of hers which is, one could say, controlled entirely by her addiction rather than by her self.

Let us now turn our attention to Gary Watson's compatibilist account of free will. Remember that I have chosen the accounts of Frankfurt and Watson because they are arguably the most influential compatibilists currently working on the free will debate and because they exemplify the compatibilist strategy of locating free will in structural features of the will.

For Watson the relevant answer to the questions, 'What makes us agents?' and 'What makes us responsible to one another?' is, like Frankfurt, the capacity for critical reflection, which he terms 'normative intelligence'. He says:

> We are agents because (and insofar as) we shape our lives by the exercise of normative intelligence; we are answerable to interpersonal norms of criticism because our lives are (in part) reflections of this capacity.[39]

The capacity for critical reflection is intricately tied, according to Watson, to our being reason-responsive. Watson defends the view that to be free means to be able to

[38] Frankfurt, H. 'The Faintest Passion', pp. 101-102.
[39] Watson, G. *Agency and Answerability*, p. 2.

do or get what one wants. He notes that this conception of freedom has traditionally been accused of being unable to explain cases of unfree action, such as those of a kleptomaniac, because it is typically assumed that being able to do or get what one wants conflates free action with intentional action, and the kleptomaniac certainly still acts intentionally. The example of the kleptomaniac is often cited in the free will debate as a prime example of a character who cannot overcome an irresistible impulse. This case is notably quite similar to that of the unwilling addict provided by Frankfurt.

To overcome the objection that his conception of freedom cannot explain cases of unfree action, Watson draws the distinction between wanting and valuing. Opting for a Platonic understanding of practical reasoning, Watson argues that reason and desire are independent sources of reasons for action. Reason determines what has value and, being oriented towards 'the Good', will always value worthy states of affairs. Since judging some state of affairs as valuable implies that one desires the promotion of this state of affairs, reason provides motivation for action, indeed for Watson reason should be seen as the "original spring of action".[40] Desire, on the other hand, is, Watson argues, non-rational and its objects, contra those of reason, may not necessarily be thought of as valuable.

Watson centres his account of free agency on 1) the notion of a free agent's being able to critically evaluate and judge the value of possible alternate courses of action, which he calls our valuational system and 2) those desires that finally motivate us to action, which he calls our motivational system.[41] According to Watson, a free agent is one whose actions flow from his valuational system and an action is most freely performed when the agent both desires and values the same course of action. He says:

> The problem of free action arises because what one desires may not be what one values, and what one most values may not be what one is finally moved to get... If there are sources of motivation independent of the agent's values, then it is possible that sometimes he is motivated to do things he does not deem worth doing. This possibility is the basis for the principal problem of free action: a person may be obstructed by his own will.[42]

It is possible to act 'unfreely', according to Watson, because it is possible for one's valuational and motivational systems to not correspond. According to Watson, this is

[40] Watson, G. 'Free Agency', p. 17.
[41] Ibid, p. 25.
[42] Ibid, pp. 18-23.

precisely the case with the oft-cited kleptomaniac. Although the kleptomaniac might be motivated to steal, Watson argues that she could not reasonably judge stealing, all things considered, to be valuable. The kleptomaniac, Watson argues, can certainly be seen as acting intentionally but not as acting freely.[43] He says:

> The contention will be that, in the case of actions that are unfree, the agent is unable to get what he most wants, or *values*, and this inability is due to his own "motivational system." In this case the obstruction to the action that he most wants to do is his own will… the agent is obstructed in and by the very performance of the action.[44]

Recall that both Frankfurt and Watson locate free will in structural features of the will. An important feature of the will for both is our capacity to engage in critical reflection and to control which of our values or second-order volitions ultimately move us to action. Watson's agent is in control, and thereby free, when she acts in accordance with her valuational system. She is most free when she both desires and values the same object. Frankfurt's agent exercises freedom of the will to the extent that she is able to secure her will to her second-order volitions; she is free when she acts on her endorsed desires. Under Frankfurt's account, we, in an important sense, take responsibility for our desires when we identify with them and endorse them as truly our own, as constituting what we really want. Under Watson's account, we are morally responsible for those actions that we perform because we really want to perform them. We are morally responsible when our actions flow voluntarily from our valuational and motivational systems. As already mentioned, both accounts posit a deeper self which critically reflects – a self which endorses desires or evaluates possible courses of action.

Although I believe that Watson and Frankfurt's accounts of agency are among our most influential and sophisticated accounts, their accounts, and compatibilism generally, fail in thinking that by providing us with free will, or free action, they have justified our belief in desert-entailing moral responsibility. However, in this, I will argue, they are mistaken. Following the incompatibilist tradition, I argue that in order to be morally responsible in the desert-entailing sense we must be self-determining, and self-determination, as the compatibilist concedes, cannot be made

[43] Given Watson's claim that the kleptomaniac does not do what she most wants Watson is committed to the view that the rational self is the true self. Frankfurt is also committed to this view.
[44] Watson, G. 'Free Agency', p. 15. Importantly, Watson argues that one's valuational and motivational systems must, for the most part, correspond. If a person was never motivated by her values we would deny, he argues, that these were her values at all.

commensurable with determinism. Although Frankfurt and Watson's accounts provide us with a picture of self-control – and this, I believe, is the cornerstone of our best accounts of free will and agency – in moving the goal posts from self-determination to self-control they are, I will argue, no longer able to justify the type of moral responsibility which drives the free will debate. To see why this is the case we need to examine the distinction provided by Watson between attributability and accountability.

1.5 Gary Watson: Accountability and Attributability

In a recent article entitled 'Two Faces of Responsibility'[45] Watson argues that there are two crucially distinct types of responsibility both of which have distinct ethical import. These are the weaker attributability and the stronger accountability. Watson argues that while both attributability and accountability involve a sense of moral blame, the blame associated with attributability is evaluative rather than normative. By evaluative blame Watson means that when we attribute blame to someone we do not see this type of blame as necessarily entailing the further response of punishment. Punishment is only warranted, according to Watson, when we hold someone accountable for their actions – accountability being the further added perspective of responsibility which, given its relationship with punishment, seemingly maps onto the traditional concept of desert-entailing moral responsibility. Attributability, then, is, for Watson, a basic sense of responsibility which we have in virtue of being the types of agents we are, whereas to argue that we are accountable for our actions may indeed involve showing something more.

Let us examine both types of responsibility in more detail. Let us start with attributability; what it means to be responsible in the sense that our actions are attributable to us. From what Watson calls 'the aretaic perspective', we are responsible for the ends we pursue – our goals and intentions. This type of responsibility has to do with the life we lead, what we find important, believe, value, desire and so on. Our actions, because they, for the most part, flow voluntarily from our desires, for example, are attributable to us as their 'authors'. Attributability is therefore appraisal of the agent as an intentional being. According to Watson:

[45] Watson, G. *Agency and Answerability*, pp. 260-288.

> Attributability has an importance to ethical life that is distinct from concerns about accountability. Responsibility is important to issues about what it is to lead a life, indeed about what it is to have a life in the biographical sense, and about the quality and character of that life. These issues reflect one face of responsibility…Concerns about accountability reflect another.[46]

Attributability, Watson argues, is the "core notion of responsibility with which the self-disclosure view is concerned",[47] views such as his and Frankfurt's which see an agent as identifying with or endorsing one of her values or second-order desires and in so doing making it truly 'hers' and taking responsibility for having it. Watson argues that self-disclosure views:

> are prompted by a concern with agency and attributability rather than with control and accountability. The significant relation between behaviour and the "real self" is not (just) causal but *executive* and *expressive*. When thought or behaviour are exercises of what Dewey calls an agents moral capacity, they and their results are open to distinctive kinds of evaluation. These evaluations are inescapably evaluations of the agent because the conduct in question expresses the agent's own evaluative commitments… This brings out the way in which aretaic appraisal involves an attribution of responsibility. To adopt an end, to commit oneself to a conception of value in this way, is a way of taking responsibility. To stand for something is to take a stand, to be ready to stand up for, to defend, to affirm, to answer for. Hence one notion of responsibility – *responsibility as attributability* – belongs to the very notion of practical identity.[48]

Accountability, the second type of responsibility, is, according to Watson, tied to the practices of reward or punishment that are traditionally linked to the notions of praise and blame, in a way that evaluations of attributability are not. Accountability, Watson argues, exactly like the traditional concept of desert-entailing moral responsibility. Because accountability is linked to our practices of reward and punishment it is essentially, or at least typically, a tripartite relationship involving at least two people and a requirement or demand, such as the demand that agents behave in socially desirable ways. When these requirements or demands are not met they usually invoke, in society, the further demand for restitution or, in the case of severe wrongdoing, retributive punishment. According to Watson:

> holding people responsible [or accountable] involves a readiness to respond to them in certain ways. To require or demand certain behaviour of an agent is to lay it down that unless the agent so behaves she will be liable to certain adverse or unwelcome treatment. Holding accountable thus involves the idea of liability to sanctions. To be entitled to make demands, then, is to be entitled to impose conditions of liability.[49]

Accountability is therefore the stronger face of responsibility since, Watson claims, viewing an action as only attributable to someone does not invoke the further

[46] Watson, G. 'Two Faces of Responsibility', pp. 263-264.
[47] Watson, G. 'Two Faces of Responsibility, p. 267.
[48] Watson, G. 'Two faces of Responsibility', pp. 270-271.
[49] Ibid, pp. 274-275.

response of restitution in the face of wrongdoing. We see this in a number of passages, for example:

> To speak of conduct as deserving of "censure," or "remonstration,"… is to suggest that some *further response* to the agent is (in principle) appropriate. It is to invoke the practices of holding people morally accountable, in which (typically) the judge (or if not the judge, other members of the moral community) is entitled (in principle) to react in various ways… Nothing in [attributability] explains or justifies any such reactive entitlement.[50]

And again:

> In one way, to blame (morally) is to attribute something to a (moral) fault in the agent; therefore, to call conduct shoddy *is* to blame the agent. But judgements of moral blameworthiness are also thought to involve the idea that agents deserve adverse treatment or "negative attitudes" in response to their faulty conduct. The former kinds of blaming and praising judgements are independent of what I am calling the practices of moral accountability. They invoke only the attributability conditions, on which certain appraisals of the individual as an agent are grounded… If we think of the aretaic perspective as concerned with the question of what activities and ways of life are most choiceworthy, then some aretaic appraisal… is what we would call moral. But even if one takes all such appraisals to be moral in a broad sense, they are independent of the particular moral norms that are invoked in accountability.[51]

Where traditionally the free will debate has focused on the stronger accountability or desert-entailing moral responsibility, Watson claims that attributability is a richly important face of responsibility since to say of someone that they are the 'author' of an action is to say something important about their responsibility for that action. He says:

> Moral accountability is only part, and not necessarily the most important part, of our idea of responsibility. The self-disclosure view describes a core notion of responsibility that is central to ethical life *and ethical appraisal*. In virtue of the capacities identified by the self-disclosure view, conduct can be attributable or imputable to an individual as its agent and is open to appraisal that is therefore appraisal of the individual as the adopter of ends. Attributability in this sense *is* a kind of responsibility. In virtue of the capacities in question, the individual is an agent in a strong sense, an author of her conduct, and is in an important sense answerable for what she does. While (strict liability aside) attributability in this sense is crucial to the practices of moral accountability, it does not all by itself underwrite them.[52]

Furthermore, while attributability is, for Watson, an unproblematic type of responsibility, accountability can give rise to scepticism about responsibility for two reasons, both of which relate to the practices associated with accountability. In order to clarify this scepticism, Watson asks two distinct questions: 1) "by what authority do we subject one another to sanctions?"[53] And 2) what form do specifically moral sanctions take? In response to the first question Watson sets out two conditions which must be met in order for any demand to be just. First, the agent must be able to satisfy

[50] Ibid, p. 265.
[51] Ibid, p. 266.
[52] Watson, G. 'Two Faces of Responsibility', p. 263.
[53] Ibid, pp. 274-275.

the demand made of her. If it is impossible for her to do so then it is unfair to sanction her for failing to do so. According to Watson, "*just* demands require the opportunity to avoid the concomitant sanction".[54] Second, just demands can only be made by a legitimate authority. To elucidate Watson provides the example of a hijacker who demands one of his captives to control the rest. Even if it is possible for her to satisfy this demand Watson claims that the demand is unjust because the hijacker is not a legitimate authority. The second of these two conditions can, Watson argues, lead to scepticism about accountability practices because supplying an adequate account of legitimate authority is highly problematic. The second question which asks about the form that specifically moral sanctions take can, according to Watson, also lead to scepticism about accountability practices – as Watson puts it: "The nature of sanctions in the moral case is not so clear".[55] If moral sanctions involve some form of disapproving attitude, such as resentment, and these attitudes invite adverse treatment then both are constitutive of a specifically moral sanction which results from blame. However, when is it fair to blame someone and sanction her based on this blame? In dealing with this question issues of avoidability are traditionally raised, and the case of whether and when it is fair to blame someone becomes cloudy. It is in light of the above considerations that Watson says:

> When one is sceptical – for one or both of these reasons – about accountability, one might be said to be sceptical about the ordinary full-fledged concept of moral responsibility. When the two perspectives are held apart… and one of them is affirmed and the other denied, the least misleading answer to the question of whether one believes in moral responsibility is: "In part yes, in part no."[56]

Here Watson seems to be saying that one can plausibly affirm the weaker, unproblematic concept of attributability while denying the stronger, highly problematic accountability. As Watson claims elsewhere, "It is no contradiction… to respond to the aretaic face of responsibility while denying the legitimacy of moral accountability".[57] This claim is pivotal to my argument. I will argue that while we should accept responsibility-as-attributability – where attributability is the only type of responsibility we in fact possess – we should deny responsibility as accountability. I argue that accountability, or desert-entailing moral responsibility, should be rejected

[54] Ibid, p. 276.
[55] Ibid.
[56] Ibid, p. 285.
[57] Ibid, p. 267.

based on the work of Galen Strawson and Thomas Nagel. I will turn to an examination of Strawson and Nagel's work shortly.

1.6 Voluntariness and Attributability & Origination and Accountability: An Important Relationship

Recall that the compatibilists believe that by giving us an account of freedom they are able to justify our belief in desert-entailing moral responsibility. Compatibilist accounts, arguably our best accounts of agency, are, however, grounded solely in voluntariness which, according to Honderich (a hard determinist), cannot provide us with desert-entailing moral responsibility. Honderich's claim that voluntariness alone cannot provide us with desert-entailing moral responsibility is further defended if we examine once again Watson's description of attributability. Recall that I am responsible for my actions in the sense that they are attributable to me as their author because they have come about as the natural result of my endorsed desires, values, intentions and so on. If I wanted to act as I acted, that is, acted not because I was coerced or compelled to act but because I actually wanted to perform the action that I performed then I can be held responsible for that action in the sense that it is attributable to me as its author. These are the very same criteria for an action being voluntary. When describing voluntariness Honderich writes:

> These are actions flowing just from embraced rather than reluctant desires, actions done in satisfying and not frustrating circumstances. Such an action… really does come from an individual – it is not against his or her desire or true nature… its initiation is a matter of *voluntariness*.[58]

Attributability, then, arises out of voluntariness alone. It seems to me, that being accountable for an action, on the other hand, requires that the action originates in me. Although Watson would want to deny this, being a compatibilist, he does himself suggest something which entails that origination is a necessary condition of accountability. He says:

> the self-disclosure view does not suffice for an understanding (or defense) of the practices of holding accountable. Arguably, control is a central issue for those practices. If so, the issue of responsibility for one's ends might reemerge in that context… [Attributability] conditions are not affected by issues about control of *one's own character*, which are driven by concerns about accountability (reward and punishment.)[59]

[58] Honderich, T. 'A Defense of Hard Determinism', p. 473.
[59] Watson, G. 'Two Faces of Responsibility', pp. 272-273 (my emphasis).

As discussed above, the claim that origination is a necessary condition for accountability is traditionally put forward by incompatibilists and is expressed succinctly by Kant, a libertarian, in the following:

> Man *himself* must make or have made himself into whatever, in a moral sense, whether good or evil, he is to become. Either condition must be an effect of his own free choice; for otherwise he could not be held responsible for it and could therefore be *morally* neither good nor evil.[60]

Recall that according to Watson I can be sanctioned for my actions if I am accountable for them, since accountability leads to a stronger sense of blame – that which typically tends to bring about punishment. Given this, accountability is precisely the same concept as desert-entailing moral responsibility. As Watson himself argues, to be sceptical about accountability amounts to being sceptical about "the ordinary full-fledged concept of moral responsibility".[61] Accountability is the stronger type of responsibility – and leads to a stronger sense of blame – precisely because to be accountable my actions must be more than merely voluntary; I must it seems be the originator of them, something which compatibilists would deny.

1.7 Rejecting Moral Responsibility: Galen Strawson and Thomas Nagel

While Watson focuses on possible scepticism associated with accountability practices Galen Strawson, in both 'The Impossibility of Moral Responsibility' and 'The Bounds of Freedom', argues that desert-entailing moral responsibility (or accountability) is impossible. The basic argument as set out by G. Strawson in these papers is the classic formulation of the incompatibilist position. According to G. Strawson:

> If one wants to think about free will and moral responsibility, consideration of some version of the Basic Argument is an overwhelmingly natural place to start… Belief in the kind of absolute moral responsibility that it shows to be impossible has for a long time been central to the Western religious, moral and cultural tradition… It is a matter of historical fact that concern about moral responsibility has been the main motor… of discussion of the issue of free will.[62]

The basic argument, crucially, does not rely on the truth of determinism, "being a priori, it holds good whether determinism is true or false".[63] I discussed the basic argument when showing libertarianism to be untenable so here I will only set it out to refresh the reader's memory:

---

[60] Kant, I. *Religion within the Limits of Reason Alone*, p. 40.
[61] Watson G, 'Two Faces of Responsibility', p. 285.
[62] Strawson, G. 'The Impossibility of Moral Responsibility', pp. 514-515.
[63] Strawson, G. 'The Bounds of Freedom', p. 441.

1. Nothing (or no one) can be *causa sui* – nothing can be the cause of itself.
2. To be truly morally responsible for one's actions one would have to be *causa sui*, at least in certain crucial mental respects.
3. Therefore, nothing (or no one) can be truly morally responsible.[64]

Compatibilists do not give the basic argument any serious attention, in fact they never actually defend their claim that desert follows from the accounts of free will that they provide – they merely believe that by giving us an account of free will or free action they have justified our belief in desert-entailing moral responsibility. If compatibilists did pay the basic argument serious attention they would attack the second premise, which claims that in order to be morally responsible one must be *causa sui* at least in certain crucial mental respects. Recall that the compatibilist grounds his account of free will in structural features of the will rather than in self-determination. For the compatibilist, then, the second premise of the basic argument is just plainly false; we are morally responsible agents because we enjoy certain levels of control over what motivates us to act. However, I argue that by exploring the work of Honderich and Watson – and in particular the relationship between origination and accountability on the one hand and voluntariness and attributability on the other – we are able to see that the compatibilist can only properly defend attributability since this is the only type of responsibility which can be justified given voluntariness alone.

Recall, that Watson's responsibility-as-attributability maps onto Honderich's concept of voluntariness and correspondingly accountability, the very same concept as desert-entailing moral responsibility, maps onto origination. If we accept that these two distinctions relate to one another in the way I have argued they do then compatibilists, since their accounts of freedom are based on voluntariness rather than origination, cannot attack the second premise after all. While compatibilists provide us with responsibility, they are not providing us with *moral* responsibility – which is after all what they purport to do.

In adopting this line of argument against the compatibilist, I join ranks with hard determinists Galen Strawson and Ted Honderich. G. Strawson directly attacks compatibilism in 'The Impossibility of Moral Responsibility' claiming that "one can have compatibilist responsibility even if the way one is is totally determined by

[64] Strawson, G. 'The Impossibility of Moral Responsibility', p. 513.

factors entirely outside one's control".[65] But, he argues, if it is the case that one is not responsible for how one ultimately is, as under the traditional compatibilist story, and it is also the case that one acts as one does because of the way one is, where this idea is succinctly expressed by the concept of voluntariness, then one cannot properly *deserve* to be punished for anything one does. G. Strawson, like Honderich, defends the claim that desert does not and cannot follow from voluntariness alone. It is for this reason that G. Strawson claims: "compatibilist responsibility famously fails to amount to any sort of true *moral* responsibility".[66]

Thomas Nagel's argument against the existence of desert-entailing moral responsibility differs from G. Strawson's argument. In contrast to G. Strawson's basic argument, Nagel's argument it relies on our experience and appeals to and extends upon our common understanding of ourselves and what it means to make a moral judgement.

Nagel begins his argument by expressing the commonly held moral intuition that when we learn that a person's action was not under his control the appropriateness of moral assessment of this action is undermined. However, according to Nagel, when we seriously examine any one of our actions there seems to be little about them that is really under our control. This, he contends, should undermine moral judgement altogether. He says:

> Where a significant aspect of what someone does depends on factors beyond his control, yet we continue to treat him in that respect as an object of moral judgement, it can be called moral luck. Such luck can be good or bad… If the condition of control is consistently applied, it threatens to erode most of the moral assessments we find it natural to make… Ultimately, nothing or almost nothing about what a person does seems to be under his control... The erosion of moral judgement emerges not as the absurd consequence of an over-simple theory, but as a natural consequence of the ordinary idea of moral assessment, when it is applied in view of a more complete and precise account of the facts.[67]

According to Nagel there are four distinct ways in which what we do is significantly beyond our control. First, Nagel speaks about 'constitutive luck' which concerns the kind of person someone is – their desires, capacities and character. According to Nagel, these seem, on the face of it, to be completely beyond our control. It is typically acknowledged that our characters are shaped by genetic predisposition and

[65] Ibid. Note how this relates to Watson's discussion of scepticism regarding accountability practices.
[66] Strawson, G. 'The Impossibility of Moral Responsibility', p. 519.
[67] Nagel, T. 'Moral Luck', pp. 530-531.

social upbringing. [68] Second, Nagel speaks about luck in the circumstances of one's life, the situations one faces and problems that one encounters – these, he claims, are clearly not under our control. Third, Nagel speaks about the relation of cause and effect, specifically the antecedent conditions and circumstances that determine how one is, and finally he speaks about the results or consequences of one's actions. It seems trivially true that the consequences of our actions are, to a large extent, beyond our control. According to Nagel, showing our actions to be so drastically out of our control in the four ways described above ought to undermine our ordinary moral assessments, particularly our assessments of moral responsibility. He says:

> If one cannot be responsible for consequences of one's acts due to factors beyond one's control, or for antecedents of one's acts that are properties of temperament not subject to one's will, or for the circumstances that pose one's moral choices, then how can one be responsible even for the stripped-down acts of the will itself, if *they* are the product of antecedent circumstances outside of the will's control? The area of genuine agency, and therefore of legitimate moral judgement, seems to shrink under this scrutiny to an extensionless point. Everything seems to result from the combined influence of factors, antecedent and posterior to action, that are not within the agent's control. Since he cannot be responsible for them, he cannot be responsible for their results.[69]

Nagel argues, then, that a serious examination of the various factors that direct or determine our behaviour strips us of responsibility. Moreover, we cannot but see that this is the case if we take seriously our place in the world. As Nagel puts it:

> something in the idea of agency is incompatible with actions being events, or people being things. But as the external determinants of what someone has done are gradually exposed, in their effect on consequences, character, and choice itself, it becomes gradually clear that actions are events and people things. Eventually nothing remains which can be ascribed to the responsible self, and we are left with nothing but a portion of the larger sequence of events, which can be deplored or celebrated, but not praised or blamed.[70]

Recall that this sentiment is reiterated by Holbach, who says:

> In whatever manner man is considered, he is connected to universal nature, and submitted to the necessary and immutable laws that she imposes on all the beings she contains… Man's life is a line that nature commands him to describe upon the surface of the earth, without his ever being able to swerve from it, even for an instant… Nevertheless, in spite of the shackles by which he is bound, it is pretended he is a free agent, or that independent of the causes by which he is moved, he determines his own will, and regulates his own condition.[71]

---

[68] Both Frankfurt and Watson would concede this point. For Frankfurt, although we can choose the will we want to have we cannot control the nature of our will simpliciter - whether our will is wholehearted or ambivalent. For Watson, the desires which we seem to just find ourselves with play a large explanatory role in his account of unfree action.

[69] Ibid, p. 535. Nagel's position here is very similar to Saul Smilansky's which I will discuss in the conclusion.

[70] Ibid.

[71] Holbach, P. 'The Illusion of Free Will', p. 462.

1.8 My Positive Account: A New Position in the Debate

If we grant that Watson's distinction between attributability and accountability maps onto Honderich's distinction between voluntariness and origination, where voluntariness by itself can only give us attributability, and further grant the soundness of Strawson's basic argument as well as the devastating implications of Nagel's work on 'moral luck' then it seems that the compatibilist, in attempting to provide us with desert-entailing moral responsibility, is on a fool's errand. Compatibilism unproblematically provides us not with desert-entailing moral responsibility or accountability but rather with attributability, which, as Watson argues, is in itself a rich and important type of responsibility.

Although Watson would certainly not be happy with his work being used in this way, given that he is a compatibilist, I believe, as mentioned above, that we can use his work in this article to dissolve the classic determinist divide. If the compatibilist concedes that all they can plausibly defend is attributability, where this is the only face of responsibility that follows coherently from voluntariness alone, then they must also concede that they are not providing us with grounds for believing in desert-entailing moral responsibility. But if the compatibilist concedes this then the classic determinist divide is dissolved since responsibility-as-attributability, and the account of agent-control that flows from this conception of responsibility, are compatible with hard determinism.

If my argument is sound and we can dissolve the classic determinist divide then I argue that we have a new position in the free will debate; a position that sees free will as grounded in structural features of the will and believes that we are responsible creatures in the sense that our actions are attributable to us, where this sense of responsibility allows us to evaluate each other as agents, and as the authors of our actions, but does not justify our retributive condemnatory practices of reward and punishment based on desert.

The account I have put forward could be objected to in two ways. First, it could be objected that the actions of my dog, Grace, are also attributable to her. But this would be to misunderstand the concept of attributability. That is, attributability stems from the idea that our actions are ours because they flow from our *endorsed* desires, values and so on. If Grace has the capacity for critical self-evaluation, the capacity to

identify with or endorse what motivates her to action, then and only then would her actions be attributable to her in the sense that has been outlined. Second, it could be objected that ultimately I am not dissolving the classic determinist divide but am rather merely defending hard determinism. This objection has clout because I argue against the compatibilist and with the hard determinist that desert-entailing moral responsibility is not compatible with determinism. Although my tendencies in the classic debate are certainly more inclined towards hard determinism, I believe that there are certain important insights offered by compatibilists that need to be defended. That is, as I have already mentioned, I believe that the accounts of agency expressed by the self-disclosure views of Watson and Frankfurt are our most advanced and promising accounts of agency and moreover certainly show us to have free will, despite the fact that these authors wrongly conclude that their views entail that we are responsible in a desert entailing sense. Contra hard determinism, then, I do not see free will as an illusion. My position is neither compatibilist nor hard determinist, but is rather a new position forged out of the two.

It may seem as though we are conceding a lot when we renounce desert. However, it seems to me that the failure of philosophers to make sense of free will as a fundamentally libertarian concept, where this is the only type of freedom which can secure desert-entailing moral responsibility, can be explained simply by realising that this type of freedom and responsibility does not exist.

CHAPTER 2: OUR PHENOMENOLOGICAL COMMITMENT

> The human commitment to participation in ordinary inter-personal relationships is… too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them; and being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question. – Peter Strawson[72]

> Even if you believe that determinism is true, and that you will in five minutes time be able to look back and say that what you did was determined, this does not seem to undermine your sense of the absoluteness and inescapability of your freedom, and of your responsibility for your choice… it remains true that as one stands there, one's freedom and true moral responsibility seem obvious and absolute to one…. The conviction that self-conscious awareness of one's situation can be a sufficient foundation of strong free will is very powerful. It runs deeper than rational argument, and it survives untouched, in the everyday conduct of life. – Galen Strawson[73]

In the previous chapter I argued that we are unable to justify our belief in desert-entailing moral responsibility at the metaphysical level; that our best accounts of free will and agency are unable to provide us with the desert-entailing moral responsibility that we experience ourselves as having and upon which our current condemnatory practices are founded. In this chapter I wish to return to the phenomenological dimension of the free will debate. More specifically, I wish to explore what is allegedly an unrenounceable and shared phenomenological commitment to experiencing (and treating) one another as morally responsible agents.

This commitment, according to Peter Strawson, is grounded in our commonplace reactive attitudes – our natural attitudinal responses to each other (and ourselves) in light of our behaviour, intentions and attitudes *qua* participants in the moral community. P. F. Strawson argues that the commonplace reactive attitudes are partly constitutive of our intra- and inter-personal relationships and the moral community itself. Our commitment to the commonplace reactive attitudes, he argues, is unrenounceable and is, by itself, sufficient justification for our belief in moral responsibility. For the sake of clarity I have briefly summarised what I take to be P. F. Strawson's main argument in 'Freedom and Resentment' below:

---

[72] Strawson, P. 'Freedom and Resentment', p. 81.
[73] Strawson, G. 'The Impossibility of Moral Responsibility', p. 516-522.

1) Human beings are always and already social creatures who engage in intra- and inter-personal relationships.
2) These relationships cause us to experience different emotions (e.g. love and anger) and different attitudes (e.g. resentment and gratitude).
3) For the most part, the attitudes and emotions we experience are commonplace – they are ordinary and natural.
4) If a person *x* is able to engage in mature inter-personal relationships and it is reasonable to demand of *x* that she behave in socially desirable ways then it is appropriate to respond to *x* with the commonplace reactive attitudes (e.g. it is appropriate or reasonable to feel resentment towards *x* if *x* acts malevolently).
5) Our commonplace reactive attitudes can be seen as expressing our beliefs (it is wrong to act malevolently and we are morally responsible for the way we act in the sense that we deserve praise and blame, and reward and punishment).
6) Our commonplace reactive attitudes are the source of free will and moral responsibility; they are the reason we are able to understand morality in the first place (our current retributive condemnatory practices are expressive of human nature).
7) The possible truth of determinism could not affect our commitment to our commonplace reactive attitudes:
    a) when we excuse or exempt *x* from moral responsibility this is not because we take determinism to be true in *x's* case,
    b) our commonplace reactive attitudes are inevitably given with the fact of human social existence,
    c) to renounce our commonplace reactive attitudes would be to drastically impoverish our lives, and the possible truth of determinism could not play a role in this decision.
8) Our commitment to our commonplace reactive attitudes is unrenounceable.

This chapter will be almost entirely devoted to P. F. Strawson's arguments in 'Freedom and Resentment' and to my claim that if P. F. Strawson is correct then there seems to exist a direct, and seemingly irresolvable, tension between our metaphysical and phenomenological commitments in the free will debate – between our best accounts of agency which deny that we are morally responsible and our shared

experiential commitment to the commonplace reactive attitudes which, for P. F. Strawson, justifies our belief in moral responsibility. However, before exploring P. F. Strawson's argument further, or the alleged tension which I believe would exist between our metaphysical and phenomenological commitments if his argument is correct, let us briefly recall that in our everyday lives we typically conceive of ourselves as genuinely free and morally responsible agents.

We typically experience ourselves, that is, as freely deciding what we want whether these decisions are mundane, such as what to have for lunch, or monumental, such as what to do with one's life. More generally, we see ourselves, to a large extent, as the makers of our own destinies and, at least most of us, hold ourselves and one another morally responsible for the way we act, praising those who do great things and punishing others because, to a large extent but not exclusively, we take them to be morally responsible for their actions.[74] Our freedom and moral responsibility are unquestioned by most, and our practices of praise and blame, and reward and punishment are founded upon these largely unquestioned beliefs. The very practices that come under pressure when we realise that we cannot metaphysically justify our belief in desert-entailing moral responsibility are ordinarily accepted because of our seemingly unchanging experience of freedom and moral responsibility. As Galen Strawson argues:

> Situations of choice occur regularly in human life. I think they lie at the heart of the experience of freedom and moral responsibility. They are the fundamental source of our inability to give up our belief in true or ultimate moral responsibility… they are the experiential rock on which the belief in true moral responsibility is founded.[75]

Even those like Galen Strawson, who is a hard determinist, argue that when we are not focusing on our metaphysical commitment it seems *prima facie* all too natural to revert back to believing in genuine free will and desert-entailing moral responsibility because, pervasively, our experience of ourselves constantly reinforces this belief. Indeed, in the quote below G. Strawson makes the stronger claim that we are incapable of holding onto a conception of ourselves that is compatible with belief in determinism. He says:

---

[74] Recall that in saying 'to a large extent but not exclusively' I am allowing for cases in which we see someone as a danger to society but do not see her as morally responsible for her actions in the desert entailing sense, as is the case with a psychologically deranged or morally undeveloped person or a sociopath.

[75] Strawson, G. 'The Impossibility of Moral Responsibility', p. 516.

> While one's attempts to grasp the consequences of determinism fully may succeed in bursts, they will in the longer term always break up on one's rock-hard commitment to a self-conception which is wholly incompatible with fully fledged, continually applied belief in determinism... A person may *theoretically* fully accept that he, or she, is wholly a product of his or her heredity and environment… and yet, in everyday life, have *nothing like* the kind of self-conception that is here required of the genuine incompatibilist determinist… such a self-conception seems scarcely possible for human beings. It seems to require the dissolution of any recognizable sense of self.[76]

It seems as though, even if we acknowledge the fact that we are merely the products of "a genetic and social roulette",[77] we carry on with life *as if* we are self-determining beings and morally responsible for our actions in the desert-entailing sense. G. Strawson's claim that a self-conception which takes seriously the possibility of determinism "seems scarcely possible for human beings [because it] requires the dissolution of any recognisable sense of self"[78] resembles a claim made by P. F. Strawson, who says that "it is useless to ask whether it would not be more rational for us to do what it is not in our nature to (be able to) do".[79] For both P. F. Strawson and G. Strawson, belief in determinism is unable to alter our phenomenologically based conception of ourselves as free and morally responsible agents. In the third and final chapters of this thesis I will argue against this view, claiming that we are indeed capable of a self-conception which not only maps onto but is also supported by our best metaphysical commitments in the free will debate as set out in chapter 1.

2.1 The Commonplace Reactive Attitudes and Moral Responsibility

> The central commonplace that I want to insist on is 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… it matters to us, whether the actions of other people – and particularly of *some* other people – reflect attitudes towards us of goodwill, affection or esteem on the one hand or contempt, indifference, or malevolence on the other... The object of these commonplaces is to try to keep before our minds something it is easy to forget when we are engaged in philosophy… what it is actually like to be involved in ordinary inter-personal relationships – Peter Strawson.[80]

Peter Strawson, in an attempt to justify our beliefs in freedom and moral responsibility (and thereby also our current retributive practices of reward and punishment) argues, in his seminal paper 'Freedom and Resentment', that the possible

[76] Strawson, P. 'Freedom and Resentment', pp. 79-81.
[77] Nielsen, K. 'Is to Understand to Forgive or at Least not the Blame', p. 252.
[78] Strawson, P. 'Freedom and Resentment', p. 81.
[79] Ibid. 'Freedom and Resentment', p. 87.
[80] Strawson, P. 'Freedom and Resentment', pp. 75-77.

truth of determinism is irrelevant to the question of whether we are free and morally responsible beings – a view that he shares with all compatibilists. Instead he argues that the source of our freedom and moral responsibility is our commonplace reactive attitudes – our natural attitudinal responses to ourselves and each other in light of our intentions and attitudes as expressed in our behaviour *qua* participants in the moral community. We need only examine our commonplace reactive attitudes, he argues, in order to justify both our freedom and moral responsibility. For P. F. Strawson the commonplace reactive attitudes include all personal feelings which arise out of our intra- and inter-personal relationships. For the sake of ease, I will talk almost exclusively about resentment and gratitude in this thesis as these two attitudes, I believe, exemplify the relationship between our attitudes, our beliefs about moral obligations, our conception of ourselves as morally responsible beings, and our practices of praise and blame and reward and punishment.

According to P. F. Strawson, our commonplace reactive attitudes are fundamentally basic to, or constitutive of, our intra- and inter-personal relationships as well as the moral community itself; they form a central part of a general system of attitudes that is "something we are given with the fact of human society".[81] In normal circumstances inter-personal relationships are largely defined by the attitudes held and emotions felt by each party towards the other. My relationship with my mother, for example, is, to a large extent, constituted by my emotional responses and attitude towards her. While P. F. Strawson acknowledges that certain attitudes may differ slightly across time or culture he maintains that "an awareness of variety of forms should not prevent is from acknowledging also that in the absence of *any* forms of these attitudes it is doubtful whether we should have anything that *we* could find intelligible as a system of human relationships, as human society".[82] According to P. F. Strawson, then, it is inconceivable that there could exist a *human* society whose members did not experience any of the reactive attitudes or feelings. Recall that for P. F. Strawson the commonplace reactive attitudes form part of a general system of attitudes which is inevitably given with the fact of human social existence and which includes all those emotions and attitudes which arise our of our intra- and inter-

[81] Ibid, p. 91 According to Strawson, then, it is inconceivable that there could exist what we would call a *human* society whose members did not experience any of the reactive attitudes or feelings.
[82] Strawson, P. 'Freedom and Resentment', p. 93.

personal relationships. The inevitability of this general system of attitudes and its being partly constitutive of our relationships entails, for P. F. Strawson, first, that our commitment to the commonplace reactive attitudes is unrenounceable since to renounce them would cause the collapse of all relationships as well as the moral community, and second, that to embark on a study of agency and free will at the level of metaphysics is to neglect the great importance we place on our reactive attitudes or to neglect the importance of the way we actually engage with each other in the world.

Since human beings, understood as moral agents, are always and already social creatures we must, by necessity, live co-operatively. We depend on one another to behave in certain socially desirable ways and for this reason, P. F. Strawson argues, we are able legitimately to place social demands on one another to behave in precisely these ways. The legitimacy of these demands makes them moral obligations. According to P. F. Strawson, if it is appropriate to place these demands on a person then she is a member of the moral community and morally responsible for her actions. He argues that if a member of the moral community breaches a moral obligation then the rest of the community will naturally feel resentment towards her, where this attitude typically expresses a desire for moral censure of the wrongdoer or in more severe cases retributive-style punishment. P. F. Strawson argues that:

> The concepts we are concerned with are those of responsibility and guilt, qualified as 'moral', on the one hand–together with that of membership of a moral community; of demand, indignation, disapprobation and condemnation, qualified as 'moral', on the other hand–together with that of punishment… [T]hese attitudes of disapprobation and indignation are precisely the correlates of the moral demand in the case where the demand is felt to be disregarded. The making of the demand *is* the proneness to such attitudes… Only by attending to this range of attitudes can we recover from the facts as we know them a sense of what we mean, i.e. *of all we mean*, when, speaking the language of morals, we speak of *desert, responsibility, guilt, condemnation, and justice*.[83]

The last sentence of this quote is crucial to understanding P. F. Strawson's main argument in 'Freedom and Resentment', namely that the commonplace reactive attitudes are the primary, or only, way we are able to make sense of morality, that they

[83] Strawson, P. 'On Freedom and Resentment', pp. 90-91 (my emphasis). Here P. F. Strawson makes the very strong claim that the only way we can understand freedom, responsibility, desert and so on is through our own reactive attitudes and the legitimate social demands which our reactive attitudes express. If we follow P. F. Strawson's argument to its logical conclusion it seems as though he is committed to the further claim that free will proper, that is strong, self-determining free will is nothing more than an illusion – something which at the metaphysical level can never be justified. Actual free will, as opposed to the mere experience of free will, must seemingly be for P. F. Strawson only an illusion, or put another way, all that free will amounts to is our conception of ourselves as morally responsible agents.

are the fundamental source of our free will and moral responsibility. The seeming inevitability and spontaneity of resentment in cases where a moral obligation has been breached entails, for P. F. Strawson, that we take ourselves to be morally responsible agents precisely because resentment expresses the belief that we are morally responsible agents. For P. F. Strawson, then, we need look no further than our own experience, or what he calls 'the facts as we know them', for proof of our unwavering commitment to the belief that we are creatures endowed with free will; the fact that we experience ourselves and each other as members of a moral community from whom goodwill can *legitimately* be demanded is enough to show that we take ourselves and each other to be free and morally responsible agents and this, for P. F. Strawson, is by itself enough to justify our beliefs in freedom and moral responsibility (as well as the retributive condemnatory practices that are founded on these beliefs).

Ishtiyaque Haji writes:

> In Strawson's view, the question about the conditions under which an agent is morally responsible is identified with the question of the conditions under which it is *appropriate* to hold an agent morally responsible. These conditions, in turn, are explained in terms of susceptibility to *reactive attitudes.*[84]

P. F. Strawson argues that it is only reasonable or legitimate to place demands on a person if she is deemed a mature member of the moral community, as morally responsible for her actions. The source of her moral responsibility is our responding to her (as well as her responding to herself and others) with the commonplace reactive attitudes. As is just alluded to, P. F. Strawson divides the commonplace reactive attitudes into three categories: the personal, moral and self reactive attitudes, where these are responses to another's behaviour towards the self, another's behaviour towards another, and one's own behaviour towards another respectively. Since it is seemingly inevitable and natural for us to respond with the commonplace reactive attitudes in our relationships with each other we tend to leave unquestioned the beliefs in free will and moral responsibility which these attitudes express. Rather, as P. F. Strawson argues, these beliefs seem completely justified in light of the commonplace reactive attitudes.

The commonplace reactive attitudes are to be contrasted with the objective attitude which will be defined and explored below. The distinction between the objective and commonplace reactive attitudes is fundamental to P. F. Strawson's three

[84] Haji, I. 'Compatibilist Views of Freedom and Responsibility', p. 204 (my emphasis).

arguments against the relevance of determinism to our free will and moral responsibility. These arguments will be critically examined in chapter 3.

2.2 The Objective Attitude

P. F. Strawson puts forward the objective attitude as an answer to the question: 'under what conditions are the commonplace reactive attitudes *not* natural, reasonable or appropriate?' In response to this question he discusses two distinct sets of considerations which undermine, either partially or completely, the commonplace reactive attitudes. The first set of considerations involves situations in which it is reasonable to claim that an agent acted accidentally or through coercion. Although situations of this sort undermine the reactive attitudes – because we do not hold agents morally responsible for accidental or coerced behaviour – they do so only for the accidental or coerced action itself. Situations of this sort invite us to view this one action, and not an agent, as the inappropriate object of the reactive attitudes. The agent remains one to whom the commonplace reactive attitudes are ordinarily natural, reasonable and appropriate because under normal circumstances she is a mature and fully-functioning member of the moral community. When her actions are not coerced or accidental then it is appropriate to hold her morally responsible and to respond towards her with the commonplace reactive attitudes.

The second set of considerations involves situations which invite the adoption of the objective attitude. The objective attitude being that which is typically directed at those we do not see as fully, or at all, morally responsible, at those who it is not reasonable, or legitimate, to place social demands upon since they do not engage in fully-fledged, mature intra- and inter-personal relationships.[85] According to P. F. Strawson:

> To adopt the objective attitude to another human being is to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained… Seeing someone, then as warped or deranged or compulsive in behaviour or peculiarly unfortunate in his formative circumstances – seeing someone so tends, at least to some extent, to set him apart from normal participant reactive attitudes on the part of one who sees him, tends to promote, at least in the civilised, objective attitudes.[86]

[85] 'Typically' because P. F. Strawson argues that we can adopt the objective attitude as a strategy in situations of strained involvement. It is also partially adopted towards children and patients in psychoanalytic treatment.
[86] Strawson, P. 'Freedom and Resentment', p. 79.

Those towards whom we would clearly adopt the objective attitude towards include, for example, the psychologically deranged and morally undeveloped. For the sake of clarity imagine a sociopath who murders his neighbour. According to P. F. Strawson's definition of the objective attitude, although society would ordinarily respond to this crime with the retributive reactive attitudes of resentment, this attitudes is inappropriate in this case because the subject is not seen as fully, or at all, morally responsible – as an agent on whom social demands can legitimately be placed. Rather, the subject is seen as "something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained".[87] Although the subject in this case would still be imprisoned, he would not be imprisoned because we see him as deserving such punishment in the sense of being morally responsible for the murder; he is imprisoned because, being psychologically deranged, he is a danger to society, a danger that must be managed.[88] It is in this vein that P. F. Strawson claims:

> To the extent to which the agent is seen in this light he is not seen as one on whom demands and expectations lie in that particular way in which we think of them as lying when we speak of moral obligation; he is not, to that extent, seen as a morally responsible agent, as a term of moral relationships, as a member of the moral community.[89]

2.3 The Commonplace Reactive Attitudes and the general thesis of determinism

Since moral responsibility, for P. F. Strawson, amounts to being the appropriate object of the participant reactive attitudes and lacking moral responsibility amounts to being the appropriate object of the objective attitude it seems as though if determinism were true – which would entail that we are not morally responsible agents – then rationality would prescribe that we adopt the objective attitudes across the board to all persons at all times. P. F. Strawson formulates this concern as follows:

> What effect would, or should, the acceptance of the truth of a general thesis of determinism have upon the reactive attitudes? More specifically, would, or should, the acceptance of the truth of the thesis lead to the decay or the repudiation of all such attitudes? Would, or should, it mean the end of gratitude, resentment, and forgiveness; of all reciprocated adult loves; of all the essentially *personal* antagonisms?[90]

---

[87] Ibid.

[88] Compare this sanction with the blame associated with responsibility-as-attributability (discussed in the previous chapter). In the case I have discussed the murder is attributable to the agent; the murder belongs to him in the sense that he is the author of it. He is not, however, accountable for the murder in the strong desert-entailing sense.

[89] Strawson, P. 'Freedom and Resentment', p. 86.

[90] Strawson, P. 'Freedom and Resentment', p. 80. In this quote we clearly see P. F. Strawson including all attitudinal and emotional responses that arise out of personal relationships under the class of commonplace reactive attitude

P. F. Strawson answers 'no' to all of the above questions. He presents three arguments in support of the claim that determinism would be unable to undermine our commitment to the commonplace reactive attitudes.[91] (In an examination of P. F. Strawson's 'Freedom and Resentment' R. Jay Wallace, in his book *Responsibility and the Moral Sentiments*,[92] labels these arguments (1) the internal argument, (2) the naturalist argument and (3) the pragmatic argument. For the sake of clarity and ease I will use Wallace's labels throughout the rest of my discussion of 'Freedom and Resentment').

1. The internal argument states that whether or not the demand for goodwill is satisfied is not affected by determinism. That is, what we are doing when we excuse or exempt *x* from moral responsibility is not the result of our believing determinism to be true. Our response to the satisfaction or dissatisfaction of the demand for goodwill hinges not on whether *x* was determined to act as *x* acted but rather on whether *x* acted intentionally, where even the proponent of determinism would not claim that intentional action is undermined by determinism. [93]

2. The naturalist argument states that the reactive attitudes are inevitably given with the fact of human social existence, and that determinism, whether it is true or not, cannot affect our spontaneously experiencing these natural attitudes and feelings.

3. The pragmatic argument states that our lives would be dramatically impoverished without the reactive attitudes, and the possible truth of determinism cannot play a role in our decision to so impoverish our lives. Given that the reactive attitudes are partly constitutive of our intra- and inter-personal relationships a life without them would also mean a life without such relationships. Wide support for these arguments can be found among compatibilists. Consider the following quotation from Susan Wolf's paper, 'The Importance of Free Will': "a world in which human relationships are restricted to those that can be supported in the absence of the reactive

[91] Once again, it seems that there is a crucial difference between being committed to something and the actual existence of that something. It seems possible, in some sense, to be committed to an illusion.
[92] Wallace, R. Jay. *Responsibility and the Moral Sentiments.* 1996.
[93] Wallace writes: "if [a bodily movement] was not intentional, it will generally not express any particular choice that the agent has made, and so it will not provide grounds for thinking that a moral obligation we hold the agent to has been violated" (*Responsibility and the Moral Sentiments*, 133).

attitudes is a world of human isolation so cold and dreary that any but the most cynical must shudder at the idea of it".[94]

According to P. F. Strawson, then, determinism could not undermine our commitment to the commonplace reactive attitudes because (1) it does not affect what we are doing when we hold someone morally responsible, (2) it does not and cannot affect our responding with the reactive attitudes since these are inevitably given with the fact of human social existence and (3) it cannot bear upon the question of our lives being drastically impoverished should we choose to give up the reactive attitudes. Recall that for P. F. Strawson, what we are actually capable of doing given the types of creatures we are far outweighs what it would seemingly be more rational to do in the face of determinism. He says:

> The human commitment to participation in ordinary inter-personal relationships is… too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no longer any such things as inter-personal relationships as we normally understand them; and being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question… A sustained objectivity of inter-personal attitude, and the human isolation which that would entail, does not seem to be something of which human beings would be capable, even if some general truth were a theoretical ground for it… we cannot, as we are, seriously envisage ourselves adopting a thoroughgoing objectivity of attitude to others as a result of theoretical conviction of the truth of determinism; and… when we do in fact adopt such an attitude in a particular case, our doing so is not the consequence of a theoretical conviction which might be expressed as 'Determinism in this case', but is a consequence of our abandoning, for different reasons in different cases, the ordinary inter-personal attitudes.[95]

If these three arguments are correct then it does seem to be the case that determinism is ineffectual against our commitment to the commonplace reactive attitudes. Indeed, if P. F. Strawson is correct and the commonplace reactive attitudes are constitutive of our intra- and inter-personal relationships, and the moral community itself, then it seems as though our commitment to them must be unrenounceable. However, if this is the case then there exists a confict, or tension, between our metaphysical commitments on the one hand, which deny that we are morally responsible agents and our phenomenological commitment to the commonplace reactive attitudes on the other which justifies our belief in genuine free will and moral responsibility. G. Strawson notes the force of this allegedly

[94] Wolf, S. 'The Importance of Free Will', p. 106. Those relationships that could be sustained in the absence of the reactive attitudes would seemingly be ones based solely on the objective attitude or stance.
[95] Strawson, P. 'Freedom and Resentment', pp. 81-82 (my emphasis).

irresolvable tension and describes it as "a very real conflict of commitment".[96] He claims that:

> While we have a deep and perhaps inderacinable commitment to the reactive attitudes and practices, it is also in our nature to take determinism to pose a serious problem for our notions of responsibility and freedom… our commitments are complex and conflict.[97]

In the following chapter I aim to dissolve this tension. I will argue, contra P. F. Strawson, that not all of our so-called commonplace reactive attitudes are constitutive of our relationships and the moral community in the way he argues. I claim that we are able to renounce, and in fact ought to renounce, the specifically retributive reactive attitudes. These attitudes, I claim, are not necessary for the maintenance of the moral community, in fact it seems as though they significantly undermine the morality behind our moral practices. Given this, I argue that our practically basic commitment to the commonplace reactive attitudes is not in fact unrenounceable and the tension between our metaphysical and phenomenological commitments is in the end resolvable.

[96] Strawson, G. 'On "Freedom and Resentment"', p. 72.
[97] Ibid, p. 71.

CHAPTER 3: DISSOLVING THE TENSION

In this chapter I will argue that while we may for the most part have the impression that we are unrenounceably committed to experiencing ourselves and one another as morally responsible beings, this phenomenological commitment is in fact renounceable. Contra P. F. Strawson, I hope to show that only some of our so-called commonplace reactive attitudes are indeed commonplace (and hence unrenounceable) and that, in particular, the retributive reactive attitudes, such as resentment, can be renounced without this renunciation causing the collapse of either our relationships or the moral community; and crucially, without this renunciation challenging the idea that we have the type of freedom and responsibility-as-attributability outlined in chapter one.

Recall that in chapter two I claimed that if P. F. Strawson is correct in 'Freedom and Resentment' then there exists a seemingly irresolvable tension between our best accounts of free will on the one hand and our conception of ourselves as morally responsible beings on the other. The seemingly irresolvable nature of this tension hinges on the three arguments briefly set out in chapter two, namely the internal, naturalist and pragmatic arguments as labelled by R Jay Wallace. Recall that these arguments are put forward by P. F. Strawson as responses to the threat which he believes determinism presents to our inter-personal relationships and the moral community itself. Recall that he says:

> What effect would, or should, the acceptance of the truth of a general thesis of determinism have upon the reactive attitudes? More specifically, would, or should, the acceptance of the truth of the thesis lead to the decay or the repudiation of all such attitudes? Would, or should, it mean the end of gratitude, resentment, and forgiveness; of all reciprocated adult loves; of all the essentially *personal* antagonisms?[98]

In this chapter I will critically examine the three arguments which P. F. Strawson presents as responses to the above questions, and, citing the work of R. Jay Wallace, Pedro Tabensky and Galen Strawson, will argue that none of them is able to seal the reactive attitudes off from the so-called threat of determinism.[99] Given that these three

[98] Strawson, P. 'Freedom and Resentment', p. 80. In this quote we clearly see P. F. Strawson including all attitudinal and emotional responses that arise out of personal relationships under the class of commonplace reactive attitude

[99] Remember that I am not endorsing determinism as true. For our purposes it does not matter whether determinism is true or not. At the metaphysical level it is impossible to justify self-determination and

arguments fail to do the work that P. F. Strawson wants them to I argue that we must examine the fairness of our current moral practices in light of the incompatibilist's concerns with justifying desert-entailing moral responsibility. This examination, I argue, ought to result in the renunciation of certain of the reactive attitudes, namely the retributive reactive attitudes. Importantly, then, the possible truth of determinism, I argue, can and must result in the renunciation of certain of our reactive attitudes, and contra P. F. Strawson, I argue that this renunciation will not mean the end of "all the essentially *personal* antagonisms".[100]

3.1 The 'Internal' Argument

First, let us examine the internal argument. (Recall that I will be using Wallace's labels of P. F. Strawson's arguments throughout my discussion for the sake of ease and clarity). The internal argument states that what we are actually doing when we hold people morally responsible is unaffected by the general thesis of determinism; that our response to the satisfaction or dissatisfaction of the demand for goodwill is unaffected by the general thesis of determinism. There seems to me to be a grain of truth in this argument. I agree with P. F. Strawson that the truth of determinism does not factor in our *ordinary* practice of holding people responsible. When we argue that an agent is not morally responsible this is not typically due to the universal truth of determinism, it is because the agent has acted under coercion, or accidentally, or because the agent is psychologically deranged or morally undeveloped. In this, P. F. Strawson is, I believe, perfectly correct. R. Jay Wallace, too, spends much time arguing for this in *Responsibility and the Moral Sentiments*. Wallace argues that determinism does not and cannot undermine our ascriptions of moral responsibility in the same way that excuses or exemptions undermine such ascriptions. If an agent is excused or exempted she is not held morally responsible.[101] Determinism, Wallace argues, cannot be seen as the generalisation of either an excuse or an exemption and cannot therefore do any normative work where our ordinary practices of holding

---

this is precisely the type of freedom we would need to have in order to be morally responsible agents. It is for this reason that we may as well *assume* determinism.

[100] Strawson, P. 'Freedom and Resentment', p. 80.

[101] Note that excuses relate to P. F. Strawson's first class of considerations – accident or coercion – that locally undermine or repudiate the participant reactive attitudes. And exemptions relate to the second class of considerations – psychological derangement or moral undevelopment – which invite the adoption of the objective attitude or stance.

responsible are concerned. Both P. F. Strawson and R. Jay Wallace argue that our ordinary moral practices should serve as the starting point for any examination of freedom and moral responsibility. Our practices, both argue, express our beliefs about moral responsibility, our moral practices, that is, imply our beliefs. Moreover, our practices are grounded in our natural responses to one another and to wrongdoing. P. F. Strawson writes:

> What *is* wrong is to forget that these practices, and their reception, the reactions to them, really *are* expressions of our moral attitudes and not merely devices we calculatingly employ for regulative purposes. Our practices do not merely exploit our natures, they express them.[102]

Likewise for R. Jay Wallace who writes:

> To make sense of the debate, we need an interpretation of the facts about responsibility that makes them dependent – in the right ways – on our practice of holding people responsible… the debate should be seen as essentially a normative debate, about the conditions that render it appropriate to hold a person morally responsible… we make best sense of [the] facts by interpreting them nonmetaphysically, taking them to be bound up with our practice of holding people responsible.[103]

For both P. F. Strawson and R Jay Wallace the possible truth of determinism does not affect what we are saying about *x* when we hold her morally responsible for an action or excuse or exempt her from moral responsibility. When we hold *x* morally responsible or excuse *x* from moral responsibility, our doing either is based on the *appropriateness* of holding *x* morally responsible and never on the thought that determinism applies in the case of *x*.

However, there is a big difference between this question and the further question, 'ought the thesis of determinism to affect what we are doing when we hold *x* morally responsible?' I argue that the possible truth of determinism ought to affect the way in which we respond to those who we deem it appropriate to hold morally responsible.

The distinction between those we hold responsible and those we do not is indeed an important one and should be held onto, however it cannot do the work of justifying our current retributive condemnatory practices. I have already argued that we are only responsible in the sense that our actions are attributable to us as their authors. Given that we cannot metaphysically justify our belief in desert-entailing moral responsibility we must reassess the fairness of condemnatory moral practices, such as

---

[102] Strawson, P. 'Freedom and Resentment', p. 93.
[103] Wallace, R. Jay. *Responsibility and the Moral Sentiments*, pp. 85-95.

retributivism, which are grounded in our deserving reward and punishment. Since we cannot justify desert we must reject retributive condemnatory practices; determinism must play a role in assessing the fairness of the way we respond to those who we hold responsible and those who we do not.

3.2 The 'Naturalist' Argument

What Wallace terms P. F. Strawson's naturalist argument states that the reactive attitudes and emotions are inevitably given with the fact of human social existence. We can begin to understand what P. F. Strawson means if we recall his claim that the commonplace reactive attitudes are natural responses to the legitimate social demands which we make of one another. Put another way, when we engage in relationships with one another and place expectations on one another – which is inevitable given that we are social and necessarily interdependent beings – we open ourselves up to being satisfied or disappointed. The various emotions and attitudes we experience as a result of our inevitable engagement in relationships are themselves inevitable, spontaneous and natural. Given this, P. F. Strawson argues that it does not matter whether determinism is true or not, belief in determinism is powerless to change the way we respond to one another precisely because the reactive attitudes are inevitable.

This argument has *prima facie* intuitive force because first, the claim that human beings are always and already social creatures seems plainly true and second, we can easily grasp the idea that the reactive attitudes accompany our intra- and inter-personal relationships. However, this argument hinges on P. F. Strawson's claim that the commonplace reactive attitudes include all of the essentially personal attitudes and emotions. R Jay Wallace, in *Responsibility and the Moral Sentiments'* objects to this broad classification of the reactive attitudes. According to Wallace, the broadness of this classification entails that the reactive attitudes no longer have any recognisable connection to our beliefs about wrongdoing. It is not so easy to see how love, for instance, is connected with these beliefs. P. F. Strawson claims that our reactive attitudes are expressions of legitimate demands, however according to Wallace he does not follow this claim far enough. If our reactive attitudes are indeed expressive of the demand for goodwill then, Wallace argues, they must be essentially connected with the belief that a moral obligation has, or has not, been breached. However, it

seems as though this connection becomes unclear if we accept P. F. Strawson's broad classification of the reactive attitudes as including all of our personal attitudes and emotions.

For Wallace, the stance of holding *x* responsible is characterised by blame and the moral sanctions, where blame precisely is the susceptibility to retributive reactive attitudes (such as resentment) and the moral sanctions serve to express these reactive attitudes.[104] The reactive attitudes, Wallace argues, must include only resentment, indignation and guilt and must be held separate from the moral sentiments which would include emotional responses to other moral values. Wallace argues that narrowing the class of the reactive attitudes down to just these three allows them to hang together properly as a class because we can see clearly the connection between resentment, indignation and guilt, the demand for goodwill and the belief that a moral obligation has been breached. To quote Wallace rather lengthily:

> Strawson takes the reactive attitudes to include the full range of feelings we are susceptible to in virtue of participating with people in interpersonal relationships … It seems to me a mistake, however, to interpret the reactive emotions in this encompassing manner… The inclusive interpretation of the reactive attitudes frustrates any attempt to provide an informative account of what unifies this set of emotions as a class. More specifically, on this approach it becomes extremely difficult to characterise the reactive attitudes as having distinctive propositional objects… Holding someone responsible comes to be interpreted as a susceptibility to feelings that have no privileged connection with beliefs about the person who is responsible. But this seems false, given the characteristic focus of the stance of holding people responsible. That stance is essentially a disposition to respond in certain ways to the moral wrongs that people commit, but we can make sense of this defining connection between the stance and moral wrongs only if we suppose that the reactive emotions in terms of which the stance is understood have their own propositional object.[105]

If Wallace is correct and the reactive emotions are to include only resentment, indignation and guilt then the claim that the reactive attitudes are inevitably given with the fact of human social existence loses it *prima facie* appeal precisely because it does not seem obvious that our being involved in relationships simpliciter necessarily entails susceptibility to the retributive reactive attitudes and emotions. As Wallace puts it:

> The reactive attitudes are not coextensive with the emotions one feels toward people with whom one has interpersonal relationships, rather they constitute a particular category of emotions specially distinguished by its constitutive connection with expectations. If this is right, however, then it may not be as difficult as Strawson suggests to picture human life

[104] It is through the inevitability of the reactive attitudes and their connection with our beliefs about moral obligations and wrongdoing that P. F. Strawson aims to defend and justify our retributive condemnatory practices. These, he argues, express our nature as human beings and without them we would lack the basic resources to recognise people as people.
[105] Wallace, R. Jay. *Responsibility and the Moral Sentiments*, pp. 10-31.

> without the reactive attitudes. Even if interpersonal relationships are inevitable for humans, it does not follow that the reactive attitudes are similarly inevitable; there may be cultures whose members do not have in their repertoire the quasi-evaluative stance of holding people to expectations in the way that is connected with resentment, indignation, and guilt.[106]

Pedro Tabensky and Galen Strawson both discuss cultures whose members do not respond to one another, or themselves, in these especially retributive ways. While G. Strawson explores the culture of Buddhist monks who often do not respond with the participant reactive attitudes at all having "succeeded in altering quite profoundly through meditative practices their experience of themselves (and others) as acting, thinking, and feeling beings",[107] Tabensky cites the Triqui people of Oaxaca who do not respond with the retributive reactive attitudes in the face of moral transgression at all.[108] Wallace, too, explores the conceptual possibility of what he calls 'a shame culture'. He says:

> There seems to be nothing in the very idea of the reactive emotions that rules out the possibility of shame cultures. The distinctive features of resentment, indignation, and guilt do not seem to be given along with the bare fact of human social life, and so one can see how there might be human communities whose members are not subject to these emotions. Second, a shame culture would not necessarily be one in which there are no recognisable ethical norms, construed as norms that make social cooperation possible, nor would it be the case that the members of such a culture would not internalise these norms, in the sense of having incentives for compliance with the norms whose effectiveness is potentially independent of externally administered sanctions and rewards.[109]

The cases discussed by Tabensky and G. Strawson offer support to Wallace's objection against the naturalist argument. While P. F. Strawson argues that the reactive attitudes are inevitably given with the fact of human social existence, neither the Triqui people nor Buddhist monks respond to one another with the retributive reactive attitudes. In fact both cultures seem to exemplify the conceptual possibility of a 'shame culture' put forward by Wallace. Contra P. F. Strawson we certainly see both cultures as human societies and yet neither, it seems, exhibit resentment, indignation or guilt: the specifically retributive reactive attitudes which, according to P. F. Strawson, spontaneously and naturally express our disapproval of a subject who violates a moral obligation.

---

[106] Ibid.
[107] Strawson, G., 'On "Freedom and Resentment"', p. 99.
[108] Tabensky, P. 'Transfiguring Judgment Practices', p. 143.
[109] Wallace, R. Jay. *Responsibility and the Moral Sentiments*, pp. 38-39. Note that Wallace's shame culture fits nicely with responsibility-as-attributability as discussed by Watson and endorsed by my metaphysical account of freedom, agency and responsibility as set out in section 1.

Wallace's objection to P. F. Strawson's broad classification of the reactive attitudes seems to me to be a very strong challenge to the naturalist argument. Furthermore, Tabensky and G. Strawson's counterexamples to P. F. Strawson support Wallace's conceptual possibility of a 'shame culture' whose members do not respond to one another with the retributive reactive attitudes, showing that these attitudes are certainly not inevitably given with the fact of human social existence. The naturalist argument, then, seems fraught with complications; there are far too many questions to be asked about the reactive attitudes and their inevitable connection with human society for this argument, by itself, to be successful in showing the threat of determinism to be impotent.

3.3 The 'Pragmatic' Argument

Recall that the third argument, the pragmatic argument, states that our lives would be drastically impoverished were we to renounce the reactive attitudes.[110] And for P. F. Strawson this would seemingly be the rational thing to do if determinism were true. Put another way, if determinism were true it would seemingly follow that it would be more rational to adopt the objective attitude towards one another at all times because the truth of determinism would mean that we were not morally responsible beings. But, P. F. Strawson argues, to adopt the objective attitude at all times would impoverish our lives so dramatically that it is practically inconceivable that such an attitude could be sustained. According to P. F. Strawson, "it is useless to ask whether it would not be more rational for us to do what it is not in our nature to (be able to) do".[111] Recall Wolf's horror at the thought of a world without the reactive attitudes: "a world in which human relationships are restricted to those that can be supported in the absence of the reactive attitudes is a world of human isolation so cold and dreary that any but the most cynical must shudder at the idea of it".[112]

However, the adoption of the objective attitude certainly does not follow from acceptance of a position which denies that we are morally responsible beings, such as the position I defended in chapter one. It is far too quick to argue, as P. F. Strawson

[110] Remember that for P. F. Strawson the reactive attitudes include all emotional responses we have in virtue of engaging in intra- and inter-personal relationships.
[111] Strawson, P., 'Freedom and Resentment', p. 87.
[112] Wolf, S. 'The Importance of Free Will', p. 106.

does, that the reactive attitudes stand and fall together so that we either adopt the commonplace reactive attitudes or the objective stance towards one another. If we focus once again on Wallace's objection to P. F. Strawson's broad classification of the reactive attitudes and agree with Wallace that the class of reactive attitudes should only include resentment, indignation and guilt then P. F. Strawson's move seems even quicker – to argue that as soon as we renounce the reactive attitudes (which if Wallace is correct include merely the retributive reactive attitudes) we are necessarily adopting the objective attitude towards one another seems almost impossible to swallow. The examples provided by Tabensky of the Triqui people of Oaxaca, by G. Strawson of Buddhist monks, and the conceptual possibility of a 'shame culture' put forward by R. Jay Wallace clearly show that communities can exist without certain of P. F. Strawson's reactive attitudes. Contra P. F. Strawson, then, it seems that there are certain attitudes and emotions that we can easily renounce without this either entailing that we adopt the objective stance towards one another at all times or that there would no longer exist interpersonal relationships as we know them. In short, renouncing the retributive reactive attitudes as the result of belief in determinism is not only possible but moreover would not entail the drastic impoverishment of our lives. Contra P. F. Strawson, then, it seems that renouncing the retributive reactive attitudes does not necessarily entail the adoption of the objective stance, and importantly it seems as though a world which did renounce the arguably archaic retributive reactive attitudes would be far less impoverished than the world we live in today. I will defend this claim further in the next chapter where I examine what would in fact follow from adopting a position, like my own, which denies that we are morally responsible beings.

In brief, it does seem that our ordinary practices of holding responsible are typically unaffected by determinism – what we are doing when we hold someone morally responsible (or conversely excuse or exempt someone from moral responsibility) is not decided by the universal truth of determinism. However, the inevitability of the reactive attitudes given the fact of human social existence, and the impoverishment of our lives if we gave up the reactive attitudes both seem largely questionable if not plainly false.

3.4 The Fairness of our Moral Practices

The incompatibilist in the free will debate may assume for the sake of argument that our practice of holding one another morally responsible is a given, and may then question the fairness of these practices. This would be for the incompatibilist to examine the fairness of our moral practices internally to the practices themselves. It is perhaps for this reason that Wallace claims: "a successful answer to incompatibilist worries must directly engage the issue of fairness".[113]

Wallace discusses two separate but compatible dilemmas regarding our practice of holding people responsible that could result from the truth of determinism: the practical dilemma and the dilemma of rationality. Recall that incompatibilists argue that strong freedom of will is necessary for desert-entailing moral responsibility; that without strong freedom of the will it is unfair and inappropriate to hold people morally responsible in the desert-entailing sense. Wallace writes:

> Of course, if the incompatibilist is right about this, and if determinism is true, then it will turn out that a condition for the fairness of holding individuals morally responsible is never satisfied. Would it follow from this that the whole practice of holding people morally responsible is irrational, unwarranted, or unreasonable? Perhaps – but not because we were initially seeking an external, rational justification for the process, but because ordinary moral scrutiny reveals the practice to be unjustified, so that there is a moral objection to be lodged against a distinctively moral activity. Moreover, this conclusion would remain disconcerting even if, as [P. F.] Strawson maintains, we have no choice but to adopt the stance of holding people morally accountable. In that event we would find ourselves necessarily committed to a moral stance that cannot satisfy our own standards of moral justifiability – a kind of practical dilemma… Being caught up in the practice of holding people morally responsible, and also committed to moral norms of fairness, we might well be led to the conclusion that the practice is essentially unfair, and this conclusion would remain an important and troubling one, even it would not lead us to cease holding people responsible.[114]

Wallace introduces the dilemma of rationality, the second dilemma which could result from determinism while discussing what he calls P. F. Strawson's pragmatic argument. Recall that this argument states that belief in determinism could not affect the rationality of our choosing to drastically impoverish our lives by renouncing the reactive attitudes. According to Wallace:

> rationality would seem to pull us in two different directions at once: pragmatic considerations, concerning the gains and losses of our activities for human life, would give us reason to retain the practice of holding people morally responsible regardless of whether determinism is true, but moral considerations of fairness would give us reason not to hold people responsible if determinism should be true. This would be a dilemma of rationality…

[113] Wallace, R. Jay. *Responsibility and the Moral Sentiments,* p. 96.
[114] Ibid, pp. 98-99.

> if the traditional incompatibilist arguments lead to a dilemma of this sort, that is reason enough to taken them to be a source of concern.[115]

Recall that for Wallace, as for P. F. Strawson, our current moral practices should be our starting point when we are thinking about free will and moral responsibility. For both philosophers, if we were to renounce our current moral practices we would lack the resources to recognise people as people. It is because of this that, in the quote above, Wallace claims that pragmatic reasons would lead us to hold onto our practices of holding people morally responsible regardless of whether determinism was true. It is the idea that determinism would make these practices unfair that causes Wallace to take the incompatibilists concerns seriously.

Galen Strawson in his response to his father, 'On "Freedom and Resentment"', illustrates clearly the typical incompatibilist concern. G. Strawson writes:

> Although our thoughts about determinism appear in actual fact quite impotent to disturb our natural and unconsidered reactive attitudes and feelings… it also seems very difficult for us not to acknowledge that the truth of determinism or of non-self-determinability brings the propriety of the reactive attitudes seriously into doubt… Defenders of the reactive attitudes may be unwise to seek to strengthen their position by appealing to the fact that commitment to the reactive attitudes is, unlike the opposed commitment, practically basic. For the incompatibilist 'pessimists' may then reply that, while the commitment they are concerned to stress is of an essentially more theoretical character, it appears to represent the simple truth. There is a very real conflict of commitment... It is in our nature to be deeply committed to the reactive attitudes. But it is also in our nature to take determinism to pose a serious threat for the notions of freedom and responsibility.[116]

According to G. Strawson the worry of the hard determinist is not shown to be groundless by an appeal to our 'practically basic' commitment to the participant reactive attitudes. Even if we are in fact committed to the reactive attitudes this does not, he argues, provide the right sort of justification for our beliefs in freedom and moral responsibility. Taking ourselves to be free and genuinely morally responsible agents is not the same as actually being free and genuinely morally responsible agents. If we cannot justify the type of freedom required for genuine desert-entailing moral responsibility then we cannot, according to G. Strawson, hope to justify moral practices that are founded on such freedom.

The practical and rational dilemmas which determinism poses concerning the fairness of our retributive condemnatory practices should by themselves cause us to question these practices. These dilemmas taken together with the fact that it indeed seems possible to renounce the retributive reactive attitudes without causing the

[115] Ibid, pp. 102-103.
[116] Strawson, G. 'On "Freedom and Resentment"', p. 72.

collapse of the moral community ought to result in our renunciation of our retributive practices. While both Wallace and P. F. Strawson claim that our current condemnatory practices are crucial to the formation and maintenance of the moral community and the moral self, Wallace, by providing us with the conceptual possibility of a 'shame culture', has shown that this may not be the case.[117] Recall that Wallace claims that such a culture could uphold ethical norms without the threat of retributive-style punishment and reward. Given this, there seems little reason to argue that retributive condemnatory practices are necessary for the maintenance of the moral community or indeed for the formation of the moral self. The examples provided by Tabensky of the Triqui people and G. Strawson of Buddhist monks not only support the above claim but also show that it is certainly possible to see oneself and others as responsible agents without holding oneself and others accountable in the desert-entailing sense. I can just as easily respect my own and others' agency by acknowledging that they are the author of their actions in the sense that these are attributable to them – the type of responsibility which I defended in chapter one.

In this chapter I have shown that the tension between our metaphysical and phenomenological commitments can be dissolved. I argued that our seemingly unrenounceable, phenomenological commitment to treating one another as morally responsible beings can in fact be renounced. First, I argued that the internal argument is flawed. While it is true that what we are ordinarily doing when we ascribe, excuse or exempt someone from moral responsibility is not typically affected by determinism, determinism ought to affect our moral practices at the more basic level of assessing the fairness of our moral practices. Second, I argued that the naturalist argument is defeated by Wallace's objection to P. F. Strawson's broad classification of the reactive attitudes. Moreover I showed how Wallace's objection is supported by Tabensky and G. Strawson's counterexamples to the supposed inevitability of the retributive reactive attitudes. Third, I argued that the pragmatic argument fails since to renounce the retributive reactive attitudes is not to renounce all of the moral sentiments, in which case renouncing the retributive attitudes would not drastically impoverish our lives. Finally, I argued that the stronger claim that we ought to retain our retributive condemnatory practices because these are necessary to the formation

[117] Wallace, R. Jay. *Responsibility and the Moral Sentiments,* pp. 38-39.

and maintenance of the moral community and moral self is questionable given the conceptual possibility of Wallace's 'shame culture' where ethical norms are upheld and complied with without the threat of retributive punishment.

Given the above, it does not seem as though our phenomenological commitment to experiencing and treating ourselves and one another as genuinely free and morally responsible agents is in fact unrenounceable. We could, and should, strive to renounce the retributive reactive attitudes, as well as our retributive condemnatory practices because these are based on the unjustified belief in desert-entailing moral responsibility. Our commitment to truth and reason ought to trump our practically basic, phenomenological commitment – a commitment which it now seems can in any event be renounced.

The tension which seems to exist between our metaphysical and phenomenological commitments can be dissolved if we focus on truth and reason, as well as the very powerful dilemmas which the possibility of determinism raises. Although we may still experience ourselves as free and morally responsible agents, given the pervasive existence of choice in our lives, we must constantly remind ourselves that taking ourselves to be genuinely free and morally responsible agents is not the same as actually being genuinely free and morally responsible agents, and in light of this knowledge we must examine, reflect on and criticise our beliefs about ourselves as well as the practices founded on these beliefs. In the next chapter I will turn my attention to such an examination after briefly exploring two other non-standard positions in the free will debate.

CHAPTER 4: THE RAMIFICATIONS

> We can make sense of the notion of autonomy or self-determination on the compatibilist level but, if there is no libertarian free will, no one can be ultimately in control, ultimately responsible, for this self and its determinations… If people lack libertarian free will, their identity and actions flow from circumstances beyond their control… Being the sort of person one is and having the desires and beliefs one has, are ultimately something one cannot control, which cannot be one's fault; it is one's luck. And one's life, and everything one does, is an unfolding of this… One chooses and acts, but this follows from who one is, ultimately as a "given". Any factor for which one is appreciated, praised, or even loved is ultimately one's luck – Saul Smilansky[118]

This chapter has two very modest aims. First, it aims to show that there is already a movement towards breaking away from the traditional positions in the free will debate and second it aims to show that once we renounce our retributive attitudes and practices the future is not at all as bleak as P. F. Strawson and Wolf would like us to believe. In this chapter, then, I will briefly explore two other non-standard positions in the free will debate, namely the positions advocated by Saul Smilansky and Derk Pereboom, indicating where my position is similar to or differs from their positions. The second aim of the chapter flows out of my discussion of Pereboom who, drawing on the work of Ted Honderich, explores the actual ramifications of adopting a position which denies that we are morally responsible beings.

Saul Smilansky, in his paper, 'Free Will, Fundamental Dualism, and the Centrality of Illusion'[119], argues as I do that we should be committed to a position in the free will debate which brings together certain elements from both the compatibilist and hard determinist positions. Smilansky argues that while it seems that we can only be committed to one or other of these positions this assumption of monism, as he terms it, is far from conceptually necessary. According to Smilansky both compatibilism and hard determinism are by themselves flawed positions, but, he argues, we can take from each position what works and combine these insights to form "a mixed, intermediate position".[120] He says:

> compatibilism and incompatibilism are indeed logically inconsistent, but it is possible to hold a mixed, intermediate position that is not fully consistent with either… a true "economy of intuitions" cannot *afford* to sacrifice the strength of either our compatibilist or

[118] Smilansky, S. 'Free Will, Fundamental Dualism and the Centrality of Illusion'
[119] Kane, R (ed.) *The Oxford Handbook of Free Will* (2002) pp. 489-505.
[120] Smilansky, S. 'Free Will, Fundamental Dualism and the Centrality of Illusion', p. 491.

> incompatibilist instincts… The initially counterintuitive step of rejecting the Assumption of Monism thus allows us to proceed along a new path that ultimately runs closer to the intuitive field than do either of the conventional monisms.[121]

According to Smilansky, compatibilism has two closely related flaws: first it only provides us, relative to the libertarian, with a shallow account of autonomy and because of this is guilty of ignoring the ultimate unfairness of treating people as if they deserve reward and punishment when they do not possess the type of free will necessary to warrant such treatment. Hard determinism is flawed, he claims, because it fails to respect persons as purposive agents. Here he draws on the oft-cited difference between a kleptomaniac and an ordinary thief. According to Smilansky, the hard determinist's failure to take into account the difference between these two characters is a very large problem with their account of agency. To elucidate he says:

> That hard determinists are indifferent to such distinctions and ethical imperatives is morally outrageous… One chooses and acts, but this follows from who one is, ultimately as a "given". Any factor for which one is appreciated, praised, or even loved is ultimately one's luck. That compatibilists are indifferent to such ultimate arbitrariness, shallowness and injustice is morally outrageous… the point I would most like to stress is that we need to try out new ways of combining [compatibilism and hard determinism]… it is not that we are *missing* something in order to appreciate that either the compatibilist or the hard determinist perspective is, in the end, the true one. Rather, to be entirely blind to the virtues of *either* of these two perspectives is to fail to see the case on free will.[122]

Smilansky's fundamental dualism is based, then, on the partial validity of both compatibilism and hard determinism as well as the partial inadequacy of both. This part of Smilansky's project is very similar to my own. I agree with Smilansky, then, with respect to denying the assumption of Monism; like Smilansky I believe that there are important insights in both compatibilism and hard determinism which need to be rescued from each and combined to form a new position. Although I agree with Smilansky's assessment of the flaws of compatibilism I am unsure that hard determinism is conceptually incapable of drawing subtle distinctions between a kleptomaniac and an ordinary thief. Like the compatibilist, the hard determinism is able to show that these two characters have different reasons and thereby different causes explaining their behaviour, and being able to provide these different explanations is to be able to respect persons as purposive agents.

I find the second part of Smilansky's project quite contentious. Smilansky claims that we are fortunately delusional in that we believe that we are genuinely free and

[121] Smilansky, S. 'Free Will, Fundamental Dualism and the Centrality of Illusion', pp. 491-492.
[122] Ibid, p. 497.

morally responsible beings. He argues that these illusory beliefs play a positive role in our moral lives and should be maintained. He argues that people should not be made aware of the fact that, contrary their beliefs about themselves, they are not genuinely free or morally responsible. If people were made aware of the truth of the human condition then, he argues, moral chaos would ensue. He says:

> humanity is fortunately deceived on the free will issue, and this seems to be a condition of civilised morality and personal value… people not under illusion would have great difficulty in functioning.[123]

However, given Tabensky and G. Strawson's counterexamples discussed in the previous chapter, as well as the conceptual possibility of Wallace's 'shame culture', believing that we are genuinely free and morally responsible beings does not seem to be necessary to the continued functioning of the moral community or to our sense of personal value. Since I believe that retribution is not a necessary feature of our lives, I also do not think that the illusion of moral responsibility is a necessary illusion. The idea that free will and moral responsibility are merely illusions is something that Smilansky shares with hard determinists, so here I part ways with both Smilansky and hard determinism generally.

Derk Pereboom, in his paper 'Living Without Free Will: The Case for Hard Incompatibilism',[124] argues that life without genuine, libertarian free will and desert-entailing moral responsibility would not in fact be devastating to our sense of meaning and purpose, to our sense of self or to our personal relationships. Here Pereboom draws on Honderich's discussion of life hopes from his paper 'A Defense of Hard Determinism'[125]. In this paper Honderich argues that there are two sets of feelings and attitudes that we can adopt towards our life-hopes, personal relationships and so on. The first set is justified if and only if we are the types of beings for whom the future is not fixed but open, if we are the originators of our actions. This is a future in which determinism is not true. The second set is justified if our actions flow voluntarily from the types of people we are. As Honderich puts it:

> These are actions flowing just from embraced rather than reluctant desires, actions done in satisfying and not frustrating circumstances. Such an action… really does come from an

[123] Ibid, p. 500-502.
[124] Kane, R (ed.) *The Oxford Handbook of Free Will* (2002), pp. 477-488.
[125] Feinberg, J & Shafer-Landau, R (eds.). *Reason and Responsibility, Eleventh Edition.* (2002), pp. 467-480.

> individual – it is not against his or her desire or true nature… its initiation is a matter of *voluntariness*.[126]

Honderich argues that we can have life hopes based on voluntariness alone. If I am unemployed I can still hope that in the future this will change; determinism being true need not undermine or in any way affect this kind of life-hope. Similarly Pereboom argues that:

> Given that we lack knowledge of how our futures will turn out, we can still reasonably hope for success in achieving what we want most even if we turn out to be creatures of our environment and our dispositions.

But what, he asks, becomes of the idea that we are able to achieve our goals and are praiseworthy for doing so? Here Pereboom considers the formation of moral character. It is typically accepted that our moral characters are formed by the upbringing we receive from our parents. But this does not prevent people from feeling proud if they have an excellent moral character. According to Pereboom, seeing our achievements along similar deterministic lines should not lead us to feel dismay. He says, "most people are capable of facing the truth without incurring much loss, and those for whom it would be painful will typically have the psychological resources to cope with the new understanding".[127]

We can see our personal relationships along similar lines; we can see people as the originators of their actions or as exhibiting voluntariness alone. According to Pereboom, if determinism is true then we must reject all personal feelings which are only justified if we are the originators of our actions, such as our retributive feelings of resentment. Here Pereboom argues, contra P. F. Strawson, that if we did renounce these reactive attitudes in light of our commitment to the truth this would not entail either that we should adopt the objective stance or that it would be appropriate to do so if determinism were true. Recall that P. F. Strawson argues that the possible truth of determinism would entail that we must adopt the objective attitude towards one another at all times which would mean the end of all interpersonal relationships, something which given our social natures it is practically impossible to conceive. Contra P. F. Strawson, Pereboom says:

> In my conception, some of the reactive attitudes would in fact be undermined by hard determinism… For some of them, such as indignation, presuppose that the person who is the object of the attitude is morally responsible. I claim, however, that the reactive attitudes that we would want to retain either are not threatened by hard [determinism] in this way or

[126] Honderich, T. 'A Defense of Hard Determinism', p. 473.
[127] Pereboom, D. 'Living Without Free Will: The Case for Hard Incompatibilism', p. 482.

> else have analogues or aspects that would not have false presuppositions. The complex of attitudes that would survive by no means amounts to Strawson's objectivity, and they would be sufficient to sustain good relationships.[128]

To elucidate, Pereboom explores indignation, forgiveness, guilt and repentance, gratitude and mature love. Indignation is unjustifiable if determinism is true but there are other attitudes, such as hurt feelings and distress, which, Pereboom argues, play the same communicative role as indignation and are perfectly consistent with voluntariness alone. According to Pereboom the impossibility of desert-entailing moral responsibility means that we must no longer see forgiveness as the willingness to overlook deserved blame and punishment. Forgiveness can be accommodated if we understand it as merely the "willingness to cease to regard past immoral behaviour as a reason to dissolve or weaken a relationship".[129] Guilt and repentance, which seem necessary for the maintenance of good relationships and moral integrity, do indeed seem threatened by the fact that we are not morally responsible agents. However, it is not the case that we need to believe ourselves to be morally responsible in order to know that we have done something morally wrong and to feel remorse and the desire to refrain from similar behaviour in the future. Gratitude may seem equally threatened by the absence of moral responsibility, however being thankful to someone for their kindness does not entail that they must be morally responsible for being kind. Gratitude is perfectly compatible with voluntariness alone because of the fact that voluntary actions flow from desires which an individual embraces. Finally mature adult love may seem endangered if we do not have libertarian free will and desert-entailing moral responsibility, but this is surely not the case. According to Pereboom:

> Moral character and action are loveable whether or not they merit praise. Love of another involves, fundamentally, wishing for the other's good, taking on her aims and desires, and a desire to be together with her. [The possible truth of determinism] threatens none of this.[130]

Positions like my own, Smilansky's and Pereboom's:

> [Endanger] neither relationships with others nor personal integrity. [They] might well jeopardize certain attitudes that typically have a role in these domains. Indignation and guilt would likely be theoretically irrational… But such attitudes are either not essential to good relationships, or they have analogues that could play the same role they typically have. Moreover love – the reactive attitude most essential to good personal relationships – is not clearly threatened … at all.[131]

---

[128] Ibid, pp. 483-484.
[129] Ibid, p. 484.
[130] Ibid, p. 486.
[131] Ibid, p. 487.

Rather, adopting a position like my own, Smilansky's or Pereboom's – positions which deny that we are genuinely free agents in the sense that we deserve to be praised and blamed and rewarded and punished for our actions – is beneficial. As Pereboom explains, human beings experience moral anger in the face of wrongdoing because we believe that we are morally responsible beings, and this anger very easily turns into the desire for retributive punishment, torture and eventually the death penalty. Adopting a position which rejects this type of moral responsibility questions the rationality of moral anger and hopefully ought to curtail the expression of anger that results in retributive urges, torture, and death.[132] Some would argue that it is only because of moral anger that we resist oppression, injustice and abuse however believing that someone has done wrong and believing that they are morally responsible for this wrong are two distinct beliefs, and the former should be enough of a reason to continue resisting oppression, injustice and abuse.

According to Pereboom, retributivism is "one of the most naturally compelling ways for justifying criminal punishment".[133] However, it must be rejected since it is founded on the unjustified belief that people deserve to be rewarded for morally good and punished for morally bad behaviour. To argue that there is a practical reason for continuing to treat one another as morally responsible agents is to condone morally unfair treatment. We still need to have recourse to action in the face of wrongdoing but this recourse should not be retributive in nature. According to Pereboom, we should adopt, what he terms, the quarantine model. He argues that while we do not blame or punish a child sick with the Ebola virus we do quarantine her because she presents a danger to society. He argues that we should respond to wrongdoing similarly. When a person commits a crime we should respond to them as a potential, or actual, danger to society and measure our response by the severity of the crime. If, for example, a person has shoplifted Pereboom recommends monitoring her, if she commits a murder a more severe form of quarantine is clearly advisable. This type of response to wrongdoing is not only justifiable but also, Pereboom argues, sidesteps the oft cited concerns that Utilitarian deterrence measures would at times justify punishing the innocent, prescribing punishment that is unduly severe or treating people as a means to an end, all of which are morally unacceptable. Adopting the

[132] Ibid, p. 488.
[133] Pereboom, D. 'Living Without Free Will: The Case for Hard Incompatibilism', p. 479.

quarantine model allows us to deal with a potential or actual danger to society with the aim of reforming or rehabilitating the wrongdoer without undermining our metaphysical commitment. Moreover, we remain, under this model, perfectly able to morally educate; the role normally assigned to blame can be played by moral admonishment, which would be equally effective in bringing about moral reform. As Pereboom puts it:

> Instead of treating people as if they were deserving of blame, [we] can draw upon moral admonishment and encouragement, which presuppose only that the offender has done wrong. These methods can effectively communicate a sense of what is right and result in beneficial reform. Similarly, rather than treating oneself as blameworthy, one could admonish oneself for one's wrongdoing and resolve to avoid similar behaviour in the future.[134]

The work of Pereboom and Honderich show that adopting a position which denies that we are morally responsible beings, like my own, would not result in devastating consequences. Our sense of self, of our personal value, is not threatened by adopting such a position, and neither is the formation or maintenance of the moral self and moral community. Moral education remains possible if we adopt such a position and importantly we are still able to hold onto all those personal feelings and attitudes which P. F. Strawson argues would be lost to us.

Far from devastating our lives it seems as though giving up the retributive reactive attitudes and the retributive condemnatory practices which express these attitudes would actually be beneficial to us. We would no longer harbour moral anger which so often results in needlessly aggressive behaviour.

[134] Pereboom, D. 'Living Without Free Will: The Case for Hard Incompatibilism', p. 479.

CONCLUSION

In my thesis I have made three substantial and interrelated arguments. First, I argued that at the metaphysical level we should be committed to a new position in the free will debate which combines the insights of both compatibilism and hard determinism. The position I propose advocates a picture of agency and free will which, like Frankfurt and Watson's pictures, grounds free will in structural features of the will and denying that we are morally responsible in the desert-entailing sense advocates instead responsibility-as-attributability. Second, I argued that although we might experience ourselves as genuinely free and morally responsible agents, our phenomenological commitment to experiencing and treating one another as such is not in fact unrenounceable. Building on Wallace's work, I argued that the retributive reactive attitudes are not commonplace, that they can easily be renounced without this affecting our agency or causing the collapse of the moral community. Given that the retributive reactive attitudes can be renounced we are able to dissolve the alleged tension between our metaphysical and phenomenological commitments. Finally, I argued that we ought to focus our attention on the truth, that is, on our metaphysical commitment which denies that we are morally responsible agents in the sense of deserving reward and punishment. Following Honderich, Pereboom and Smilansky, I showed that this commitment does not undermine our personal relationships, sense of self or value, our life-hopes, projects, legal practices of moral admonishment and so on. Perhaps the most important consequence of our metaphysical commitment is that we must renounce our retributive reactive attitudes and retributive condemnatory practices – practices which express our retributive attitudes and are grounded in our unjustified belief that we are morally responsible in the desert-entailing sense. Instead, I argue with Pereboom, that fairness dictates that we should adopt something like the quarantine model – a forward-looking pragmatic form of punishment with the justified aims of deterrence, reform and rehabilitation.

BIBLIOGRAPHY

1. Ayer, A. J. 2002. 'Freedom and Necessity', in Feinberg, J. & Shafer-Landau, R (eds.) *Reason and Responsibility*, Wadsworth Group, 481-485.
2. Davidson, D. 1963.'Reasons, Actions, and Causes', in *The Journal of Philosophy*, 60(23), 685-700.
3. Frankfurt, H. 1971. 'Freedom of the Will and the Concept of a Person', *The Journal of Philosophy*, 68(1), 5-20.
4. Frankfurt, H. 1998. 'Identification and Wholeheartedness', in *The Importance of What We Care About*. Cambridge University Press. 159-176.
5. Frankfurt, H. 1998. 'The Faintest Passion', in *Necessity, Volition and Love*. Cambridge University Press. 95-107.
6. Frankfurt, H. 2002. 'Alternate Possibilities and Moral Responsibility', in Feinberg, J. & Shafer-Landau, R (eds.) *Reason and Responsibility*, Wadsworth Group, 523-528.
7. Haji, I. 2002. 'Compatibilist Views of Freedom and Responsibility', in Kane, R (ed.) *The Oxford Handbook of Free Will*, Oxford University Press, 202-228.
8. Holbach, P. 2002. 'The Illusion of Free Will', in Feinberg, J. & Shafer-Landau, R (eds.) *Reason and Responsibility*, Wadsworth Group, 462-466.
9. Honderich, T. 1993. *How Free Are You: The Determinism Problem?* Oxford University Press.
10. Honderich, T. 2002. 'A Defense of Hard Determinism', in Feinberg, J. & Shafer-Landau, R (eds.) *Reason and Responsibility*, Wadsworth Group, 467-480.
11. Honderich, T. 2002. 'Determinism as true, Both Compatibilism and Incompatibilism as false, and the real problem', in Kane, R (ed.) *The Oxford Handbook of Free Will*, Oxford University Press, 461-476.
12. Kane, R (ed.) 2002. *The Oxford Handbook of Free Will*. Oxford University Press, 3-41.
13. Kant, I. 1960. *Religion within the Limits of Reason Alone*. Trans. T. M. Greene and H. H. Hudson, New York: Harper and Row.

14. McKenna, M. 2006. ‘Where Frankfurt and Strawson meet’, in *Free Will and Moral Responsibility* (Midwest Studies in Philosophy: XXIX). French, P. A., Wettstein, H. K. & Fischer, J. M. (eds.) Blackwell Publishing: Boston MA & Oxford UK, 163-180.
15. Nagel, T. 2002. ‘Moral Luck’, in Feinberg, J. & Shafer-Landau, R (eds.) *Reason and Responsibility*, Wadsworth Group 529-537.
16. Nielsen, K. 2006. ‘Is to Understand to Forgive or at Least not the Blame’, in Tabensky, P (ed.) *Judging and Understanding: Essays on Free Will, Narrative, Meaning and the Ethical Limits of Condemnation*, Ashgate, 175-198.
17. Nietzsche, Beyond Good and Evil, translated by R. J. Hollingdale (London, Penguin, 1990) §21.
18. Pereboom, D. 2002. ‘Living Without Free Will: The Case for Hard Incompatibilism’, in Kane, R (ed.) *The Oxford Handbook of Free Will*, Oxford University Press, 477-488.
19. Smilansky, S. 2002. ‘Free Will, Fundamental Dualism and the Centrality of Illusion’, in Kane, R (ed.) *The Oxford Handbook of Free Will*, Oxford University Press, 489-505.
20. Strawson, G. 1986. *Freedom and Belief.* Clarendon University Press: Oxford.
21. Strawson, G. 1993. ‘On “Freedom and Resentment”’, in Fischer, J. M. & Ravizza, M. *Perspectives on Moral Responsibility.* Cornell University Press, 67-100.
22. Strawson, G. 2002. ‘The Bounds of Freedom’, in Kane, R (ed.) *The Oxford Handbook of Free Will*, Oxford University Press, 441-442.
23. Strawson, G. 2003. ‘The Impossibility of Moral Responsibility’, in Watson, G (ed.) *Free Will: Second Edition*, Oxford University Press, 212-228.
24. Strawson, P. 2003. ‘Freedom and Resentment’, in Watson, G (ed.) *Free Will: Second Edition*, Oxford University Press, 72-93.
25. Tabensky, P. ‘Moved Movers: Transfiguring Judgement Practices’, in Tabensky, P (ed.) *Judging and Understanding: Essays on Free Will, Narrative, Meaning and the Ethical Limits of Condemnation*, Ashgate, 129-152.

26. Wallace, R. Jay. 1998. *Responsibility and the Moral Sentiments*, Harvard University Press.
27. Watson, G. (ed.) 2003. *Free Will: Second Edition*. Oxford University Press 1-25.
28. Watson, G. 2002. 'Free Agency', in *Agency and Answerability: Selected Essays.* Clarendon Press: Oxford, 13-32.
29. Watson, G. 2002. 'Two Faces of Responsibility', in *Agency and Answerability: Selected Essays.* Clarendon Press: Oxford, 260-288.
30. Watson, G. 2004. 'Free Action and Free Will' in *Agency and Answerability: Selected Essays.* Clarendon Press: Oxford, 161-196.
31. Watson, G. *Agency and Answerability: Selected Essays.* Clarendon Press: Oxford, 1-12.
32. Wolf, S. 1993. 'The Importance of Free Will', in *Perspectives on Moral Responsibility*. J. M. Fischer & M. Ravizza (eds.) Cornell University Press, 101-118.
33. Wolf, S. 1987. 'Sanity and the metaphysics of responsibility' in *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. F. Schoeman (ed.) Cambridge University Press, 46-63.