Copyright

by

James Preston Andrew

2016

The Report Committee for James Preston Andrew
Certifies that this is the approved version of the following report:

Why No One Truly Deserves to Suffer

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Galen Strawson

Michelle Montague

Why No One Truly Deserves to Suffer

by

James Preston Andrew, B.A.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Arts

The University of Texas at Austin

May, 2016

Abstract

Why No One Truly Deserves to Suffer

James Preston Andrew, M.A.

The University of Texas at Austin, 2016

Supervisor: Galen Strawson

Suffering, as I understand it, is an intrinsically awful state in which to be. Yet, it is widely thought that suffering can be intrinsically good when experienced by someone who is guilty because the guilty deserve to suffer. In these pages, I attempt to show that commonsense morality misleads us insofar as it inclines us to think that suffering can ever be deserved independently of consequentialist considerations. I argue that the kind of responsibility required to ground one's deserving to suffer is "*ultimate* moral responsibility". Ultimate moral responsibility, I contend, should be understood as responsibility of such a kind that if one bears it for one's actions, then one is the ultimate cause of the way that one is mentally, at least in certain respects. Employing Galen Strawson's Basic Argument for the impossibility of ultimate moral responsibility, I defend the claim that *no possible being* could truly deserve to suffer. I close by defending the

Basic Argument against what I take to be three of the stronger objections that have been raised against it.

Table of Contents

Introduction

Suffering is an intrinsically awful state that, all else being equal, no one would wish to experience.[1] Yet, throughout history, and across many cultures, it has been widely believed that

(G) The guilty deserve to suffer.

But who, precisely, *are* these "guilty" who supposedly deserve to suffer? For it is evident that not *all* who knowingly act badly, or wrongly,[2] are also guilty in such a manner as to deserve suffering. Many subjects[3] - young children and dogs, for example - sometimes bring about bad states of affairs through their actions, and perhaps some can even be said to act on impermissible principles of action. But we do not think that young children and dogs are *morally responsible*[4] for their bad or wrong actions in such a manner as to

[1] I understand suffering in what I take to be the most minimalistic sense possible: as unwanted pain. Pain, by definition, hurts; however, some people apparently enjoy certain varieties of pain - for example, the pain associated with lifting weights or running a marathon. And perhaps there are some genuine masochists in the world. But when pain is unwanted, then, by definition, no one enjoys it. It is this unwanted variety of pain, the variety that is everywhere and always experienced as awful, with which I am here concerned. Of course, in the flawed world in which we find ourselves, all else is not always equal, and we sometimes choose - quite rationally, it seems - to undergo suffering in order to cure ourselves of diseases, to fulfill our obligations, to facilitate personal growth, and so on. I do not mean to deny any of this. I only mean to insist that suffering is always experienced as awful and would, therefore, never be sought after for its own sake.

[2] I take no stand here on what it means to act badly, or wrongly, for I take it that everything I say about moral responsibility holds true regardless of the moral theory to which on is committed. I do, however, presuppose a basic commitment to objectivity in morality, given that questions about moral responsibility are uninteresting at best confused at worst if there are not objective moral truths at all.

[3] For my purposes here, all that is required in order to count as a subject is being a locus of conscious experience. In other words, I understand any and all entities that it is *like something to be* to count as subjects. All agents count as subjects, though I remain noncommittal about whether the converse is true. Certainly not *all* subjects are moral agents.

[4] I take it that one is morally responsible for something x just in case one is an appropriate target of praise or blame for x.

deserve to suffer in virtue of having performed them. Indeed, the belief that subjects such as young children and dogs cannot be morally responsible for their actions has led some philosophers to draw a distinction between moral *agents*[5] and moral *patients*.[6] According to those who draw this distinction, only moral agents are morally responsible for their actions in such a manner as to be proper objects of praise and blame; moral patients may have important claims upon moral agents, but they are not themselves subject to obligations (even though their actions can, of course, be normatively evaluated). It seems to be a core assumption of commonsense morality that the average adult human is a moral agent, as evidenced by the fact that we do have well-established social practices by which we hold one another responsible for our actions in ways that we do not hold children and non-human animals responsible for their actions.

Widespread acceptance of the belief that the guilty deserve to suffer has many important practical consequences in our world. It results in people embracing retributive justifications of criminal punishment[7] that appear to be more punitive than is required for the aim of preventing future harm. It motivates some people to take matters into their own hands and mete out justice, or at least what they see as justice, where they perceive the state has failed, or will fail, to do so. It causes others to "shame" and ostracize those with whom they disagree about certain matters, not in an earnest effort to alter their adversaries' opinions, but simply in an effort to inflict some degree of psychological

[5] I understand a subject to count as a moral agent just in case the subject is capable of judging that it might be the case that there are objectively better and worse ways that he or she could act in a given circumstance.

[6] See, for example, Regan (1983).

[7] Retributive justifications for punishment hold that punishment is justified on the basis of the desert of the offender.

suffering upon their foes. It can cause feelings of anguish in the loved ones of a victim whose wrongdoer has gone unpunished.[8] And, of course, many theists throughout history have thought it just that God should cause a significant portion of humanity to suffer endless torment in hell.

Here I argue that commonsense morality leads us astray insofar as it inclines us to believe that anyone, even the guilty, can truly deserve to suffer; that is, that it can ever be intrinsically good[9] that anyone, including the guilty, should suffer. In Section 1, I identify what I take to be the three different varieties of moral desert and explain why, precisely, I believe that

(A) If one truly deserves to suffer, then one has performed at least some bad, or wrong, action for which one is responsible in the right sort of way.

In Section 2, I present two arguments that Randolph Clarke has offered in support of (G). Clarke's arguments are very modest. I present them, and explain why I believe they are unsound, not only for the sake of giving the opposition a fair hearing, but also to show that even the most modest defense of (G) (i) rests on the assumption that we are morally

[8] Consider, for example, the story that Jared Diamond tells of his own late father-in-law, Jozev Nabel, whose family was murdered by a Polish gang member during World War II. Nabel found his family's murderer and held him at gunpoint, prepared to shoot. But he decided not to pull the trigger and left it to the state to bring the man to justice. The state, however, decided to send this murderer to prison for only one year. Diamond explains that, "Until his own death, nearly sixty years after the murders of his parents and his release of his mother's killer, Jozef remained tormented by regret and guilt - guilt that he had not been able to protect his parents, and regret that he had failed in his responsibility to take vengeance." The article was by published by the New Yorker in 2008 and can be found online at: http://www.newyorker.com/magazine/2008/04/21/vengeance-is-ours.

[9] I understand something to be intrinsically good just in case it is good in itself and not only in virtue of its bringing about something *else* that is good.

responsible for our actions in a certain sort of way and (ii) does not obviously succeed even if this assumption is granted. In Section 3, I argue that the type of moral responsibility required to ground one's deserving to suffer is what I refer to as "*ultimate* moral responsibility". I define ultimate moral responsibility as responsibility of such a kind that

> (UMR) If one bears ultimate moral responsibility for one's actions, then one must be the ultimate cause of the way that one is mentally, at least in certain fundamental respects.[10]

It seems that Derek Parfit also accepts something like (UMR). Parfit also believes that we are not morally responsible for our actions in this ultimate sense and, therefore, argues for a claim that is at least *close* to the antithesis of (G):

> (U) We cannot deserve to suffer.

I present this argument in Section 4. Although I believe the argument that Parfit provides for (G) to be sound - and although it is, in fact, the very argument that first led me to believe that no one could truly deserve to suffer - I do part ways with Parfit in one important respect. For Parfit appears to believe that ultimate moral responsibility, though

[10] As we shall see at the end of Section 3, Immanuel Kant seems to have endorsed (UMR), at least in his later thinking on the topic of moral responsibility.

conceptually possible, is not metaphysically possible for us, given that our actions are (according to him) only events in time. I, however, believe that ultimate moral responsibility is conceptually impossible[11] and, therefore, that we could not be ultimately morally responsible for our actions - and thus, in turn, that we could not truly deserve suffering, even if our actions were more than events in time.

In Section 5, I construct my own argument for

(U*) No possible being could truly deserve to suffer,

which of course entails (U). My argument for (U*) relies, in turn, upon Galen Strawson's Basic Argument for the impossibility of ultimate moral responsibility. Strawson believes ultimate moral responsibility to be impossible because he holds that

(V) No one can possibly be ultimately morally responsible for the way in which they are mentally at any time.

I present Strawson's argument for (V) in Section 6. Not surprisingly, given that its conclusion is not one that many are anxious to accept, the Basic Argument has been subject to a number of criticisms. In Section 7, I defend it against what I take to be three of the stronger, and more prominent, of these criticisms, two of which are due to

[11] I believe that ultimate moral responsibility is metaphysically possible in something like the manner in which, say, round squares are metaphysically possible.

Randolph Clarke and one of which has been made separately, and somewhat differently, by both Clarke and Derk Pereboom.

1. Three Notions of Desert

It is commonly held that there are two basic types of desert: one that is forward-looking, or consequentialist, in character and one that is backward-looking, or non-consequentialist, in character. S deserves x in the consequentialist sense just in case things would go best if S were given x. I believe that there are two varieties of non-consequentialist desert. One of these is widely recognized; I refer to it as merit-based desert. Merit-based desert is grounded in one's past actions, such that S deserves x in the merit-based sense just in case S has, in the past, performed some action for which S was - and remains[12] - responsible so as to make it intrinsically good that S be given x. The other, less recognized, variety of non-consequentialist desert I call "capacity-based desert". S deserves x in the capacity-based sense just in case S has one or more capacities in virtue of which it is intrinsically good that S should receive x. I believe that we need this third variety of desert in order to make sense of cases in which we claim that it is intrinsically good that a subject should be given something despite the subject's not having earned it in virtue of his or her past actions and not in virtue of consequentialist considerations. For example, we sometimes say such things as "that infant deserves to be loved" and "that dog deserves a kind caretaker", even though it seems that infants and dogs are not the kind of beings who can earn things, given that they cannot be morally responsible for their actions. Such claims could, of course, be interpreted in a

[12] There are, plausibly, ways in which we could lose our responsibility for our past acts. If, for example, I was responsible for performing some immoral action in the past but then lost my memory of having performed the action, along with the personality traits that led me to perform the immoral act, then one might think that I no longer deserve to suffer in virtue of having performed the action.

consequentialist manner. For example, we could understand the claim "that dog deserves a kind caretaker" to express nothing other than something like "the world would contain more happiness if that dog went to a kind caretaker". However, at least in many cases, consequentialist interpretations seem unlikely to do justice to what people *intend* to express by all such claims.

I refer to both merit-based and capacity-based desert as two different varieties of *true* desert, for I believe that the everyday notion of desert is a non-consequentialist one.[13] Consequentialists who would deny this - who would insist, not only that their forward-looking conception of desert is the one we *should* use but also the one that we *do* use - would have to hold that, even by commonsense morality, content is not lost when claims of the form "S deserves x" are restated in the form "S ought to be given x". But while it seems clear that consequentialist considerations often do influence our commonsense judgments about what people ought to be given, it seems equally clear that the everyday notion of desert is not a *purely* consequentialist one. For it is commonly believed that, in many instances, it is intrinsically good that we should receive something in virtue of one or more of our past actions or in virtue of our capacities. This is why, for example, so many people view retributive justice in a favorable light: they believe that, all else being equal, justice demands that the guilty should pay for their crimes, even when exacting payment from the guilty cannot be reasonably expected to produce the best consequences. Indeed, it is why many find the expression of something morally

[13] This is of course not to say that non-consequentialist desert is the variety of desert in which we *ought* to believe, only that it is the one that is *in fact* operative in commonsense morality.

profound in the ancient legal phrase, *fiat justitia ruat caelum* ("let justice be done though the heavens fall").

To consider whether anyone could deserve to suffer in the consequentialist sense is simply to consider whether there can be cases in which the best available course of action results in some suffering. It is obvious that such cases are possible; indeed, it is an unfortunate fact about the world in which we find ourselves that such cases do not seem all that rare. And it is clear that no one could deserve to suffer in the capacity-based sense. It surely could not be the case that, for example, because one has the capacity to become a fine pianist, one deserves to suffer. It seems that one can only deserve *good* things in virtue of one's capacities (that is, assuming that one can deserve things in virtue of one's capacities at all), but one must have *earned* a bad thing in order to have deserved it. And we can only earn bad things in virtue of those of our past actions for which we are morally responsible. Thus, my discussion here is focused entirely upon the question of whether suffering can ever be *truly* deserved, which is to say that it is focused entirely upon the question of whether suffering can ever by deserved in the *merit-based* sense.

In the light of the above considerations, we can, I believe, argue as follows:

(1) Suffering is, by definition, an awful state in which to be.

(2) One could truly deserve to be in an awful state only if one is responsible in the right sort of way for having acted badly or wrongly.

Therefore,

(A) If one truly deserves to suffer, then one has performed at least some bad, or wrong, action for which one is responsible in the right sort of way.

As indicated at the beginning of this chapter, I take it for granted that suffering is intrinsically awful. This is not necessarily to say that I take it that suffering is *objectively* bad (though I do believe that it is); only that the character of the experience itself is awful, such that anyone who is suffering will wish, all else being equal, to *stop* suffering. And (2) expresses the thought that nothing at all could be truly deserved by one who is in no way responsible for anything. It seems that this is necessarily true, for, as already noted, to deserve something awful in the merit-based sense seems to require being, in some crucial respect, responsible for what one does. Thus, it seems to me that one who wishes to defend (G) should not do so by rejecting (A). A defender of (G) should concede that (A) is true and then offer some *explanation* of why it is that people sometimes truly deserve to suffer. This is precisely what Randolph Clarke does in his 2013 article, "Some theses on desert", to which I now turn.

2. Clarke's Two Arguments For (G)

Like me, Clarke believes that the everyday notion of desert is a non-consequentialist one, and he identifies two considerations that might lead us to believe that someone deserves something in the non-consequentialist sense (both of which, I believe, could ground one's deserving something in either the merit-based or the capacity-based sense).[14] First, Clarke suggests, we might invoke the notion of *fittingness*, holding that one can deserve to suffer if it is *fitting* that one should suffer. Clarke understands the question of whether it is fitting that one should suffer to be one of the factors that can go into determining "what ought to be or what is permissible".[15] Those who employ this notion of desert, and who accept that

> (GS) To feel guilty is to suffer,[16,17]

might hold that the guilty deserve to suffer on the grounds that

> (F1) It is fitting that the guilty should "feel guilty at the right time, and to the right degree", in virtue of having acted badly, or wrongly.[18]

---

[14] Clarke's notion of non-instrumental value seems to be one and the same as what I have called "intrinsic value".

[15] Clarke (2013), p. 154.

[16] One might doubt that feeling guilty actually is a form of suffering. However, guilt is a state with a negative affect; thus, whenever it is unwanted, I think that we should follow Clark in counting it as a kind of suffering.

[17] It should be noted that the lettered claims in this section are not direct quotes of Clarke but, rather, rephrased statements of some of his central claims.

[18] *Ibid*., p. 155.

And (F1) is suggestive of an alternative account of what it might mean for suffering to be *deserved*, which we may state as follows:

(F2) For someone S to deserve to suffer is for suffering by S to be fitting.

Other times, Clarke thinks, desert is construed "in terms of value".[19] According to "one view along these lines" that he considers, to say that S deserves x is to say that

(N1) It is intrinsically good[20] that x should be given to S.

People who believe that suffering is the sort of thing that can be deserved will, Clarke thinks, accept the claim that

(N2) For someone S to deserve to suffer is for it to be intrinsically good that S suffer "at the right time and to the right degree".[21]

One who understands desert in terms of value can accept the claim that

[19] *Ibid*.
[20] Clarke himself speaks of "non-instrumental" goodness, but I take it that non-instrumental goodness and intrinsic goodness amount to the same thing, and talk of the latter seems to me a bit less cumbersome.
[21] *Ibid*., p. 155.

(N3) It is intrinsically good that one who is guilty should feel guilty "at the right time and to the right degree."[22]

Clarke thinks that there can be intrinsic value in an agent's feeling guilty at the right time and to the right degree because he holds that "there is value in the recognition by one who is blameworthy for some moral offense of the fact that she is so blameworthy".[23,24]

Thus, Clarke offers us a way of making two distinct arguments for the claim that the guilty deserve to suffer. First, we can construct the *Fittingness Argument*:

(GS) To feel guilty is to suffer.

(F1) It is fitting that the guilty should feel guilty at the right time, and to the right degree, in virtue of having acted badly, or wrongly.

(F2) For someone S to deserve to suffer is for suffering by S to be fitting.

Therefore,

(G) The guilty deserve to suffer.

Second, we can construct the *Value-Based Argument*:

(GS) To feel guilty is to suffer.

[22] *Ibid*.
[23] *Ibid*.
[24] There are, plausibly, *non-moral* offenses. Clarke does not say whether he believes there to also be intrinsic value in those guilty of such offenses feeling guilty.

(N2) For someone S to deserve to suffer is for suffering by S to be intrinsically good.

(N3) It is intrinsically good that one who is guilty should feel guilty at the right time and to the right degree.

Therefore,

(G) The guilty deserve to suffer.

I believe that both the Fittingness Argument and the Intrinsic Value Argument are contestable, even setting aside questions about moral responsibility. The Fittingness Argument, in particular, rests upon what I think is an unwarranted assumption: namely, that because it is fitting that one should suffer at time t, it follows that it is also *good* that one should suffer at t. But it seems to me that we can easily imagine cases in which a subject's suffering is *fitting* but, nevertheless, *bad* and, therefore, undeserved. It is helpful here to consider the *lex talionis* principle. There is a clear sense in which it seems intuitively fitting that if A does some harmful thing, x, to B, then x should be done to A in turn. However, it is precisely because living by the principle of *lex talionis* would inevitably lead us to do barbaric things that the principle seems morally unacceptable, at least absent significant revision. But to concede this is to concede that the concept of fit is not obviously a *moral* concept at all. This is an important insight, not because Clarke relies upon *lex talionis* anywhere in his Fittingness Argument but rather because he im-

plicitly takes our intuitions about what is *fitting* to be reliable guides to important truths about what people deserve.

Clarke does not claim that the Intrinsic Value Argument, even if sound, necessarily establishes that it is always, or even generally, good that the guilty should suffer. He concedes that x's being intrinsically bad is compatible, at least on certain moral frameworks, with its also being instrumentally good in virtue of producing something else or in virtue "of its standing in some other relation to something else".[25] Conversely, x could be intrinsically good but extrinsically bad. For example, my happiness might be intrinsically good but extrinsically bad if it comes at the cost of the happiness of ten others. Thus, we could accept (2) but nevertheless think that, in fact, it is intrinsically bad for *anyone*, including the guilty, to suffer.

The Intrinsic Value Argument and the Fittingness Argument are both based upon the assumption that we are responsible for our actions in such a manner as to potentially deserve to suffer for them, even if only by feeling guilty.[26] So, the pressing question we must now answer is this: what *sort* of responsibility, precisely, must one bear in relation to one's actions and/or character in order to be in a position to *truly deserve* to suffer?

[25] *Ibid*., p. 158.
[26] Clarke himself concedes this point on page 161.

3. Why True Desert Requires Ultimate Moral Responsibility

It seems that when we judge an agent to have done something wrong or bad, what we hold the agent responsible for is his or her *decision* to act wrongly or badly. This is why intentions are of paramount importance in ethics. If an agent unknowingly causes something bad to happen, we do not hold the agent morally responsible for this bad thing (provided his or her ignorance was not culpable). So, in assessing whether the guilty can deserve to suffer, what we really want to know is whether the guilty can be responsible for their wrong or bad decisions in such a manner as to deserve to suffer for them. One might think that, in order to deserve to suffer, a guilty agent need only be the *proximate cause* of his or her bad or wrong decisions, where for x to be the proximate cause of y, x must be the primary, or the most salient, cause of y. That is, one might accept

> (PC) If some agent is the proximate cause of some bad or wrong decision, then the agent deserves to suffer.

The fundamental problem I see with (PC) is that it *overextends* the concept of moral responsibility. That is, I think that if we understand some agent S to bear responsibility of such a sort for x that S could deserve to suffer only in virtue of having been the proximate cause of x, then we will understand such responsibility to have obtained in many cases where, intuitively, it has not obtained.

To see why, it will be helpful to bring into clearer view what, precisely, is involved in being a proximate cause - and to this end, it will be helpful in turn to evaluate a case in which an agent clearly is the proximate cause of a decision to perform an action that should not be performed. Consider:

> *My Vandalism*. I decide, on my own volition, to throw a baseball through the closed window of some house simply for the pleasure of watching the glass break, knowing full well that this will involve destroying another person's property. I succeed in breaking the glass as intended.

In this case, it would certainly be reasonable to identify *me* as the proximate cause of the window's breaking. I would not have been the *only* cause of its breaking. Innumerable other causes (for example, the force of gravity, the ambient temperature, the air pressure, the Big Bang, and so on) would also have been involved, for merely throwing a baseball at a window is insufficient for ensuring that the glass will break. Nevertheless, for practical purposes, it would seem both imminently reasonable and uncontroversial to identify my decision to throw the baseball as the primary or most salient of all the causal factors that contributed to the window's breaking. Moreover, I would be swiftly and utterly dismissed by the homeowner or by law enforcement if I were to insist upon my innocence on the grounds that "Gravity did it!" or that "The Big Bang did it!".[27]

[27] When one comes to see actions as events embedded in vast causal webs then one might begin to wonder if to say that agents cause *anything at all* is really just to engage in a kind of useful fiction. For is it not the case that our actions are mere links in causal chains that reach back long before the time of

However, I believe that a common moral assumption - or, at any rate, an assumption that is *implicit* in commonsense morality, whether widely recognized or not - is sufficient for undermining (PC): namely, the assumption that

> (R1) If some agent S deserves to suffer in virtue of having done x at time t, then S must have been responsible for the way in which S was at t.

Returning to *My Vandalism 1*, it seems clear that, in order to be morally responsible for throwing the baseball through the window - let alone to be in a position to deserve to suffer for having done so[28] - it would need to have been the case that I was *responsible for the way in which I was mentally*, at least in certain crucial respects, when I decided to throw the baseball. To see why this is so, consider an alternative version of the case, in which I decide to throw a baseball through the window but, intuitively, am *not* responsible for the way that I was mentally when deciding to do so:

> *My Vandalism 2.* I have implanted in my brain an electrode by which a mad neuroscientist is afforded complete control over my mental life, which he

our own existence? And is the way in which we are mentally at a given time not completely fixed by the physical particles that constitute us - and, moreover, do these physical particles not have causal histories all of their own, of which the role they play in giving rise to our consciousness is but a small part? I am inclined to accept such a reductive variety of physicalism, but many people are not. Moreover, some reductive physicalists might nevertheless believe, or wish to argue, that we can, and do, possess the variety of responsibility required to truly deserve suffering. And although I believe that establishing the truth of a reductive account of the self would be *sufficient* for establishing that no one could truly deserve to suffer, I certainly do not think that it is *necessary* to show that such an account of the self is true in order to establish that suffering cannot be truly deserved. Therefore, for my purposes here, I set aside entirely questions about the nature of the self.

[28] In this case, presumably, I could not deserve to suffer very severely for my action.

exercised until yesterday, when he "set me free". I am unaware of ever having been manipulated by this neuroscientist, and, phenomenologically, nothing changed in my experience of the world when the neuroscientist relinquished his control over me. I now decide, on my own volition, to throw a baseball through the closed window of some house simply for the pleasure of watching the glass break, knowing full well that this will involve destroying another person's property. I succeed in breaking the glass as intended.

In this version of the case, I would not, by my intuitions, be morally responsible for the window's breaking, despite being the proximate cause of the decision that led to its breaking, given the lack of control I would have had over the formation of the psychological dispositions that turned me into the sort of person who would engage in petty vandalism. If I could not have been morally responsible for my decision to throw the baseball, then I could not be, or have been, morally responsible for the decision in such a manner as to deserve to suffer for it. To hold otherwise would be to commit ourselves to the very implausible view that one can deserve to suffer for making decisions for which one is not morally responsible.

We have thus established that

(R2) In order to be morally responsible for some state of affairs we have brought about, we must have have been morally responsible for the way in which we were

mentally, at least in certain fundamental respects,[29] when we *decided* to bring about this state of affairs.

The consequences of denying (R2) are, I believe, implausible. We clearly do not believe that for *any* decision, there is some subject who is morally responsible for it. And at least part of the story to be told about why we do not believe this is that we take it that not *all* decision-makers are morally responsible for their decisions. It seems natural to speak of very young children, those in the grip of madness, and some non-human animals as making decisions (indeed, it would seem bizarre to deny that many such subjects make decisions). However, intuitively, and according to common moral and legal practices, these subjects are not full-fledged moral agents in the sense of having obligations and being appropriate objects of praise and blame. And they are not full-fledged moral agents because we take it to be clear that *they are not really responsible for the way that they are,* mentally, in any respects at all. To be truly responsible for one's actions, one must, it seems, be the *cause* of the way that one is, mentally. We should, therefore, accept

(D1) If one is the sort of being that could deserve to suffer, then one must be *ultimately* morally responsible for the way in which one is mentally, at least in

[29] It seems that one could be fully morally responsible for what one does despite not being responsible for, say, the fact that one speaks English and not French. What ultimate moral responsibility seems to require is that we at least be responsible for possessing the beliefs and dispositions that form the basis for the decisions that we make.

certain fundamental respects.

I contend that any view on which we can be ultimately morally responsible for what we do, even when we are not responsible for the way that we are mentally, is vulnerable to a certain kind of *manipulation argument*. Manipulation arguments are usually deployed by incompatibilists[30] arguing against compatibilist[31] accounts of free will and moral responsibility.[32] I have already run a kind of manipulation argument in motivating (R2). And it seems to me that this basic form of argument can be employed to demonstrate that, in order to be morally responsible for one's bad or wrong actions in such a manner as to truly deserve to suffer for them, one would need to have been the cause of the way that one is mentally, in certain fundamental respects. For the argument seems to show that there is no variety of moral responsibility available to an ordinary agent who is the way that she is mentally because of causal factors beyond her control (for example, her genetic makeup and her previous life experiences) that is not *also* available to an agent who is the way that she is because of the doings of another agent. Consider the following case, which is essentially a more fleshed out version of *My Vandalism 2*:

Rene's Predicament. Unbeknownst to René, he is the creation, and the puppet, of

[30] Incompatibilists hold that free will and moral responsibility are incompatible with determinism - with, that is, the doctrine according to which there is at any given moment only one metaphysically possible future, given the prior states of the universe in conjunction with the laws of nature.

[31] Compatibilists hold that free will and moral responsibility are compatible with determinism.

[32] See, for instance, Kane (1996), Taylor (1974), and, perhaps most importantly, Pereboom (2001).

> a mad neuroscientist named Fred. Fred created René in a petri dish, personally composing his genome. Fred also implanted an electrode in René's brain at birth, by which Fred has determined his subject's every thought and action. Up to this moment, everything that René has done, he has done solely because it was Fred's will. But at some time t_n during René's adulthood, Fred decides to set his subject "free" and allow him to think his own thoughts and perform his own actions. Phenomenologically, René experiences t_n as just another moment in his life, so that he is never aware in any respect that he was once, but is no longer, the puppet of another agent.

Will René's choice be free from t_n onward? It seems not. Why? Because René cannot escape Fred's influence on his behavior at t_n, or at any later time because, even if Fred stops directly manipulating René at t_n, there is a crucial sense in which René can never be free of Fred. For it is because of Fred that René has the disposition, memories, hopes, desires - even the genes - that he possesses. René cannot escape Fred's influence because every decision that René makes, every action he performs, will necessarily be due to the causal interaction between his environment and the way that he is mentally. And neither René's environment nor the way that he is mentally is truly up to him, given that the way he was mentally at t_n was determined by Fred and that his environment is, by definition, not under his control. (One might feel pulled by the thought that René can play *some* role in controlling his environment insofar as he can at least choose to go one place rather

than another or associate with certain people rather than others. But any decision René makes at any given time t, including any decision about how he might alter his environment, will itself be the product of the way that he is mentally and the environmental factors influencing him at t.)

If we, like Fred, are products of influences beyond our control - if, for example, the way that we are mentally is determined entirely by some combination of environmental influences, our past experiences, our genes, and our chemical-womb environment - then our situation is not relevantly dissimilar to Fred's. Indeed, if the way in which we are mentally in certain fundamental respects is not ultimately up to us, then there is a sense in which we each have our own, personal Fred, where Fred is simply a stand-in for the myriad factors that, together, make us who we are at a given moment in time.

In order to be ultimately morally responsible for the way in which one is mentally, one would need to be more than merely a product of one's past experiences, genes, and chemical-womb environment. One would need to exist as some sort of self apart from - and, at least to some degree, immune to - these influences. And one would need to, somehow, be a product of one's own creation, such as to be the cause of one's very own existence. Immanuel Kant seems to have recognized this, writing 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

> free choice; for otherwise he could not be held responsible for it and could therefore be *morally* neither good nor evil"[33]

We should follow Kant in accepting

> (UMR) If one bears ultimate moral responsibility for one's actions, then one must be the ultimate cause of the way that one is mentally, at least in certain fundamental respects.

Derek Parfit apparently believes that ultimate moral responsibility might be *conceptually possible* in some timeless realm where agents perform actions that are more than just events in time; however, he does not believe that *we* are agents of such a sort.[34] Thus, Parfit believes that we are not ultimately morally responsible for our decisions and actions and, thus, does not believe that we can deserve to suffer. I turn now to his argument.

[33] Kant (1793/1960), p. 40.
[34] See Parfit (2011), Chapter 11.

4. Parfit's Argument

Some have taken (G), the claim that the guilty deserve to suffer, to be so obvious as to not even require defense. Arguably, Kant took this position with respect to (G), taking it as a given that, from the practical standpoint, which we must adopt when deciding how we ought to act, we take ourselves to be beings of such a kind that we can deserve to suffer in virtue of our immoral actions.[35] In Chapter 11 of *On What Matters* (vol. 1), Derek Parfit attributes the following argument to Kant:[36]

> (J) If our acts[37] were merely events in time, we could never deserve to suffer.
>
> (R) We can deserve to suffer.
>
> Therefore,
>
> (S) Our acts are not merely events in time.[38]

Parfit believes that (S) is false but concedes that Kant's argument is valid. So, in order to reject (S), Parfit must reject at least one of the argument's premises, as well. Parfit accepts (J), believing that if our acts are merely events in time, then we cannot be responsible for them in such a manner as to potentially deserve to suffer for them.[39] And

---

[35] Kant, of course, would insist that it is beyond the power of theoretical reason to ascertain the *actual* metaphysical nature of our agency.

[36] Whether Parfit is correct in attributing this argument to Kant is not of direct concern here; however, he does seem to be on firm exegetical ground.

[37] I have, up to now, been speaking of "actions". But I draw no distinction between actions and "acts".

[38] Parfit (2011), p. 216.

[39] It should be noted that many philosophers disagree with Parfit on this point, believing that free actions simply *are* events in time.

we surely could not deserve to suffer in virtue of some event's obtaining if we did not freely bring about the event. Indeed, to inflict suffering upon some person because an event that he or she did not freely bring about would seem to be the very height of injustice. So, Parfit reasons, if all of our acts just *are* events in time, then it follows that we never deserve to suffer for any of them. Thus, he rejects (R) and argues as follows:

(J) If our acts were merely events in time, we could not deserve to suffer.

(S) Our acts are merely events in time.

Therefore,

(U) We cannot deserve to suffer.[40]

I believe that (U) is both a profoundly important, and a profoundly under-acknowledged, moral truth. From the truth of (U), it follows that if suffering is intrinsically bad, then it is always, and everywhere that it occurs, bad (though as I have noted, this is compatible with its being true that there are cases in which suffering is instrumentally good[41]). Additionally, certain of our moral and legal practices appear to presuppose that the negation of (U) is true and are, therefore, in need of revision or elimination. So, important practical conclusions follow from (U) if I am correct that it is true.

---

[40] *Ibid*.

[41] For example, undergoing treatment for many forms of illness can result in suffering. But, if the treatments are effective in restoring a person to health, then the suffering they produce is good all things considered, given they are simply the means by which a good that outweighs them is produced.

Parfit's argument rests upon premises that many philosophers would reject. (J) might well be rejected by most compatibilists and even some libertarians[42] who, despite agreeing with Parfit that our actions are only events in time, nevertheless think that we can be morally responsible for our actions in important respects and might, therefore, believe that we can deserve to suffer in virtue of acting wrongly, or badly. And some agent-causalists[43] would reject (S), believing that our actions are more than events in time. But, of course, it does not follow from the fact that some philosophers would reject Parfit's premises that his argument is a bad one. And the revised argument for (U) I propose in the following section is no improvement upon Parfit's in terms of relying upon uncontroversial premises.

[42] See, for example, Kane (2003).

[43] Agent causalists believe that we exist as robust irreducible agents and possess some special causal power by which we are able to cause free actions.

5. My Alternative Argument For (U)

My concern with Parfit's argument is not that it relies upon contestable premises but, rather, that the manner in which (I) is stated at least suggests that Parfit is willing to entertain the truth of a proposition that I believe to be false: namely, that we might have deserved to suffer for our immoral actions if our actions were not only events in time. For given that the reason Parfit adduces for thinking we cannot deserve to suffer is that our actions are simply events in time, it is natural to read him as believing that (U) is only *contingently* true and, therefore, *could* have been false if our actions were *not* merely events in time. I, however, believe that (U) must be true across *all possible worlds*, for I accept the stronger claim that

(U*) *No* being could truly deserve to suffer.

I do not think that even an agent whose actions were not events in time could deserve to suffer. And I think that it is in the light of *this* reason that we ought to accept (U). For if we accepted (U) in virtue of the reason that Parfit offers - namely that the conjunction of (J) and (S) is true - we would not only believe (U) to be true for what seems to me a *bad* reason but also for a reason that is *sufficient*, but not *necessary*, for (U)'s being true. Pragmatically speaking, the unfortunate aspect of this state of affairs would be that those who *disbelieved* the conjunction of (J) and (S) could think that their disbelief gave them grounds for rejecting (U) when, in reality (at least if I am right), it

would do no such thing. Thus, insofar as it is morally important to increase the number of minds who accept important moral truths such as I believe (U) to be, it is also morally important to construct as strong of an argument for such moral truths as we can manage.

I believe that (U) is true, and should be accepted as true, not only because, like Parfit, I believe that (J) and (S) are true but also because, like Galen Strawson, I believe that

> (V) No one can possibly be ultimately morally responsible for the way in which they are mentally at any time.

and, recall, that

> (D1) If one is the sort of being that can deserve to suffer, then one must be *ultimately* morally responsible for the way in which one is mentally, at least in certain fundamental respects.

Thus, unlike Parfit, I derive the truth of (U) as follows:

> (D1) If one is the sort of being that can deserve to suffer, then one must be *ultimately* morally responsible for the way in which one is mentally, at least in certain fundamental respects.

(V) No one can possibly be ultimately morally responsible for the way in which they are mentally at any time.

Therefore,

(U*) No being can truly deserve to suffer.

Therefore,

(U) We cannot [truly] deserve to suffer.

I have already explained why I believe we must accept (D1). So, I turn now to the task of defending (V).

6. The Impossibility of Ultimate Moral Responsibility

In *Freedom and Belief*, and in subsequent writings, Galen Strawson has, I believe, established that ultimate moral responsibility is impossible. Although the most developed argument he provides for this claim is rather involved, his basic insight is simple and can be stated concisely in standard form as follows:

> (1) In order for one to be ultimately morally responsible for what one does, one would need to be ultimately morally responsible for the way in which one is mentally, at least in certain crucial respects.
>
> (2) In order to be ultimately morally responsible for the way that one is mentally, one would need to be the *ultimate cause* of the way in which one is mentally in certain crucial respects.
>
> (3) It is *impossible* to be the ultimate cause of the way that one is mentally in any respect.
>
> Therefore,
>
> (4) One cannot possibly be ultimately morally responsible for what one does.

This line of argument invites misunderstanding, so it is worth pausing at this point to clarify what, precisely, is and is not being claimed. Neither Strawson nor I deny that there is a *local* sense in which we most certainly *can* be the cause of the way that we are mentally in certain respects. It is (in part) because I once decided to learn to play tennis

that being a tennis player is now a part of who I am; it is (again, at least in part[44]) because I decided to major in philosophy as an undergraduate that I am now a philosopher. Our decisions and actions play a crucial role in shaping who we become, and so to whatever extent we are responsible for our decisions and actions, we are also responsible for the way in which we are mentally.

However, whenever we make a decision or perform an action, we do so, by definition, with a formed character, and with procedures and principles by which we make decisions, already in place. Strawson's claim, which I also believe, is that, in order to be ultimately morally responsible for, and not merely the proximate cause of, what we do, we would need to have, at some past time, freely created our own character and chosen the procedures and principles by which we make decisions. For absent such a moment of self-creation, then in the deepest metaphysical sense, we could never be truly in control of our mental life. The way that we are, our character, would, throughout all the moments of our lives, simply be *given* to us. But it is logically impossible for one to be the ultimate cause of one's character, given that one would, by definition, need to *already exist*, with a formed character and decision-making procedures and principles *already in place*, in order to truly create oneself.

[44] This qualification is necessary because even the things that we feel as though we freely decide to do are not, upon reflection, up to us to quite the extent that we might have believed, pre-reflectively. For example, while I once decided to learn how to play tennis, I did not decide not to have been born with a clubbed foot that might precluded the possibility of doing so, and I did not decide to have a father who could afford to pay for my lessons. While I decided to major in philosophy, it was by good (I am choosing to be an optimist for the moment) fortune that I happened to attend a university with a strong philosophy department, and it was largely just because it fit my class schedule that I decided to enroll in Philosophy 101 as a freshman. Life is far more contingent, and much less "up to us", than we often realize.

Strawson states this simple, two-step argument as follows:

(1) Nothing can be *causa sui* - nothing can be the cause of itself.

(2) In order to be truly morally responsible[45] for one's actions one would have to be *causa sui* at least in certain crucial mental respects.

(3) Therefore nothing can be truly morally responsible.[46]

It follows from (3) that, although there is most definitely a sense in which we can be said to make choices and distinguish actions from mere events and moral agents from moral subjects, there is nevertheless a variety of freedom - and, thus, also a variety of responsibility - that is unavailable to anyone. Given that the argument is valid, to reject (3), one would need to reject (1) and/or (2).

(1) appears to be on very firm ground. Some might wish to reject (1) on the grounds that God is a *causa sui*. But even God, it has traditionally been held, is constrained by logical possibility. And self-creation seems to be logically impossible. For in order to create oneself, one would need to exist and not exist at the same time. I am inclined to agree with Spinoza[47] that not even God could do this. But even if self-creation is possible for God, it is certainly impossible for finite beings such as ourselves.

The obvious premise to attack is (2). Many would deny that we would need to be

[45] By "truly morally responsible", Strawson means the same thing as I mean by "ultimately morally responsible".
[46] Strawson (1994), p. 5.
[47] See Spinoza (1675/1985), Pt. 1, Prop. XXXII, Corolls. I and II.

causa sui in order to be morally responsible. But recall what I argued in Section 4: to deny that one must be ultimately causally responsible for the way that one is mentally in certain crucial respects in order to be morally responsible for what one does leads to implausible consequences. So, if one wishes to deny (2), one should agree that we must be causally responsible for the way that we are mentally, in certain crucial respects, but deny that this requires being a *causa sui*. However, I do not believe that there is any logical space in which to maneuver here - though I do not take this to be self-evident, and neither does Strawson. Realizing that it takes more than two premises to convince many philosophers that their shoes are untied, he offers a much more extended version of the Basic Argument intended to motivate the idea that being a *causa sui* is indeed a prerequisite for ultimate moral responsibility.

Strawson casts this extended version of the Basic Argument in terms of free action, where "free action" is conceived as an action performed by an agent who is ultimately morally responsible for his or her actions and can truly deserve praise or blame for its performance. Thus, to act freely in the sense that Strawson has in mind is to act in a manner such that one is ultimately morally responsible for what one does. Strawson begins his argument by noting that (a) the actions for which we commonly take ourselves to bear moral responsibility are performed for a *reason*[48] and (b) such actions are ones that we perform in virtue of *how we are*; most crucially, how we are *mentally*.[49]

Strawson acknowledges that the facts about the state of one's body, one's location,

[48] It is important to note that Strawson is a reasons internalist, holding that an agent can be properly said to act for a reason so long as the agent acts on a belief and a desire. See Strawson (1986/2010), p. 24.

[49] Strawson (1994), p. 6.

and the time at which one acts play a part in determining what one does. But it is only the facts about how we are, mentally, when we act that seem to be relevant in determining whether we are morally responsible for our actions. When considering why a person has performed some action, we want to know what it was that went through the person's *mind* that gave rise to the action. Indeed, if the action was an immoral one, then what went through the person's mind when he or she performed it will be crucial in shaping our thoughts about whether, or to what extent, the person should be reprimanded for having performed the action. If, for example, a driver hits some unfortunate pedestrian crossing a street, it will be of great concern to us whether the driver *wished* to hit the person, or if, instead, the driver wished with all his might to *avoid* doing so but simply could not hit the brakes in time.

However, Strawson notes that "if one is to be truly responsible for how one acts, one must be responsible for how one is, mentally speaking - at least in certain respects".[50] The thought here is that if the responsibility-conferring aspect of one's actions is the way that one is in certain respects mentally at the moment one decides to act, then it follows that one must be responsible for the way that one is in these mental respects when one acts. For if one were not ultimately morally responsible for this, then one could not be ultimately morally responsible for the ensuing action.

And yet, the only way in which we could be ultimately morally responsible for the way that we are, mentally speaking, would be for us to have, at least in certain respects, *made* ourselves the way that we are; to have *chosen* to be as we are. But one

[50] *Ibid.*

cannot choose to make oneself the way that one is "unless one already exists, mentally speaking, already equipped with some principles of choice" to guide the decision making process.[51] We could only be ultimately morally responsible for choosing to be the way that we are, mentally, if we are "truly responsible for...having the principles of choice" by which we chose to be the way that we are.[52]

However, Strawson points out, we could only be ultimately morally responsible for having these principles of choice if we *chose* them "in a reasoned, conscious, intentional fashion".[53] And in order to have chosen our principles of choice in this fashion, we would need to *have already had principles of choice* by which to choose them. Of course, this line of reasoning leads to an infinite regress. Thus, Strawson claims that "true self-determination...requires the actual completion of an infinite series of choices of principles of choice", which is, of course, impossible.[54] The conclusion of the Basic Argument, then, is that ultimate moral responsibility is impossible, given that "it requires true self-determination".[55]

Strawson's crucial insight is this: the way in which we are mentally is, ultimately, *not up to us*. As it happens, our minds are products of some combination of our past experiences, our genetic makeup, and our prenatal womb environment. There is not a single, solitary moment in our lives during which any of these factors are under our control. But suppose it were otherwise. Suppose that we were each immaterial minds

[51] *Ibid*.
[52] *Ibid*.
[53] *Ibid*.
[54] *Ibid*., p. 7.
[55] *Ibid*.

who, at least at the very beginning of our lives as moral agents, were completely unconstrained by either our past experiences or any facts about our biology. Although many seem to be possessed of a pre-reflective intuition to the contrary, this would not in fact help matters at all. For if our immaterial minds were simply *given to us* - either by nature or by God, then we would have played no role in creating or choosing them and so could not be ultimately morally responsible for the things that we do as a consequence of having the particular soul that we do. The way that we were and the things that we did would, therefore, still ultimately be nothing more than a matter of *luck.*

It seems to me that the considerations that lead Strawson to believe ultimate moral responsibility to be impossible also render more deflationary conceptions of moral responsibility implausible, at least insofar as such conceptions are held to do all of the practical work that has traditionally been done the everyday notion of moral responsibility. For to hold that one who is *not* the ultimate cause of some action could deserve to suffer in virtue of it seems obviously immoral - if one truly comes to terms with what it means to be something less than the ultimate cause of an action. If we could adopt a God's-eye-view of the universe, proximate causes would fade away, and we would only see causes causes and their effects. The very notion of a non-ultimate, or proximate, cause of an action or an event is really just a pragmatic one, born of our own epistemic limitations. Thus, to hold that someone who is not ultimately morally responsible for what he or she does could nevertheless deserve to suffer is not to hold that someone who is *mostly* – or even *partially* – responsible for his or her actions could

deserve to suffer. It is instead to hold that someone could be responsible in such as a way as to deserve to suffer precisely for having been dealt a bad hand by the universe. If suffering by one who is, ultimately, the victim of cosmic bad luck could be deserved, then I fear that I do not understand the concept of moral desert at all.

7. Three Objections to the Basic Argument

Strawson's conclusion that the variety of moral responsibility in which we all pre-reflectively believe, and which much of our moral thinking seems to presuppose, is impossible has, as one might expect, been met with significant resistance. Indeed, Randolph Clarke perceives that "few philosophers have been persuaded by [the Basic Argument]".[56] Daniel Dennett concedes that Strawson's Basic Argument proves that *something* is impossible but believes that a robust variety of moral responsibility can nevertheless exist in the shadow of the Basic Argument's soundness.[57] Strawson himself has remarked that many have believed his argument to be "wrong, or irrelevant, or fatuous, or too rapid, or an expression of metaphysical megalomania".[58] I have already, in the previous section, explained why I believe that people like Dennett are wrong to believe that many varieties of responsibility are untouched by the Basic Argument. Although my only contention here is that no one could truly deserve to *suffer*, I hope to have also shown in Section 6 that if anyone truly deserves anything at all, this will have to be consistent with the fact that anyone ever does is, in the end, a matter of luck. It seems clear that no one could truly deserve something terrible, like suffering, in virtue of having been the victim of bad fortune, and this is all that one must see in order to agree with me that suffering could not be truly deserved - that is, if one accepts the Basic Argument as sound. So, the remainder of this chapter is devoted to defending the Basic Argument against three prominent objections that have been raised against it.

[56] Clarke (2005), p. 13.
[57] Dennett (2014).
[58] Strawson (1994), p. 8.

Objection 1: A Missing Premise?

Clarke does not believe that Strawson makes sufficiently explicit his rejection of a position defended by many of his compatibilist opponents who deny that being ultimately morally responsible for *what one does* requires being ultimately morally responsible for *the way that one is*. Clarke presents Strawson as arguing:

> (i) You do what you do, in any situation in which you find yourself, because of the way that you are.

So,

> (ii) To be truly responsible for what you do you must be truly responsible for the way that you are - at least in certain crucial mental respects.[59]

Clarke believes that, in order to make his position more explicit, Strawson should add, between (i) and (ii),

> (O) When you do what you do because of the way that you are, to be truly morally responsible for what you do, either (a) you must be truly responsible for the way that you are, at least in certain crucial mental respects, or (b) it must be

[59] Clarke (2005), p. 18.

up to you whether if you are that way, in certain crucial mental respects, then you perform that action;

and

(P) When you do what you do because of the way you are, it is not possible for it to be up to you whether if you are that way, in certain crucial mental respects, then you perform that action.[60]

Clarke worries that Strawson only engages the position of those who reject (P) insofar as "he argues for what is asserted in (P)"; he does not, Clarke thinks, offer any real defense of (O), which "some defenders of the possibility of moral responsibility appear" to reject.[61]

However, anyone who rejects (O) will be vulnerable to the sort of manipulation argument presented in Section 3 and will, therefore, have to either affirm that René acts responsibly or identify a relevant dissimilarity between his case and our own. Anyone who believes that René acts responsibly will have difficulty making sense of a number of commonly held beliefs about the nature of moral responsibility. For example, it is a part of our commonsense framework that one's moral responsibility tracks the degree to which one is in control of the way that he or she is, mentally. This is why it is commonly

[60] *Ibid*., p. 19.
[61] *Ibid*.

believed that factors such as youth (or extreme old age), the presence of a tumor in certain regions of the brain, and the ingestion of certain drugs can diminish (or eliminate) one's moral responsibility for one's actions.

Strawson's view is that, through philosophical reflection, we can come to see that moral responsibility is fundamentally impossible because, ultimately, *none* of us has any say over the way that we are at any given moment in our lives. But one who rejects (O) will have to either

> (a) tell a (seemingly *post hoc*) story about why it is that, even though we can be responsible for what we do despite not being responsible for the way that we are, there are certain causal factors, the presence of which in a causal chain leading to some action A, can undermine the agent's responsibility for A

or

> (b) hold that even subjects whom we do not generally take to be morally responsible for their actions (for example, very young children) nevertheless actually *are*.[62]

Strawson's position here seems the most plausible. We should follow him in accepting (O).

Objection 2: A Hyperbolic Conception of Responsibility?

[62] *Ibid.*

Strawson sometimes defines what I refer to as ultimate moral responsibility (he sometimes calls it "true responsibility") in terms of desert. Specifically, he has defined it as "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".[63] Strawson explains,

> As I understand 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 stress on the words 'makes sense' is important, for one certainly does not have to believe in any version of the story of heaven and hell in order to understand the notion of true moral responsibility that it is being used to illustrate. Nor does one have to believe in any version of the story of heaven and hell in order to believe in the existence of true moral responsibility. On the contrary: many atheists have believed in the existence of true moral responsibility. 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. It very clearly expresses its scope and force.[64]

[63] Strawson (1994), pp. 9-10.
[64] *Ibid*.

It should be noted that Strawson also defines ultimate moral responsibility more modestly as responsibility

> "of such a kind that it can exist if and only if punishment and reward can be fair or just without having any pragmatic justification, or indeed any justification that appeals to the notion of distributive justice."[65]

Although I understand my notion of ultimate moral responsibility to be equivalent to Strawson's, I have avoided defining it in terms of desert. I have instead, recall, defined ultimate moral responsibility as responsibility of such a kind that

> (UMR) If one bears ultimate moral responsibility for one's actions, then one must be the ultimate cause of the way that one is mentally, at least in certain fundamental respects.

But I do agree with Strawson that if one were ultimately morally responsible, then one could genuinely deserve all sorts of things - including, perhaps, eternity in heaven or hell. Moreover, I think Strawson is correct that the conception of moral responsibility that is *in fact* operative in the mind of the average person is "heaven and hell responsibility". (Whether this is in fact so is, of course, an open empirical question.) Thus, if one believes that the notion of responsibility that Strawson takes to be necessary for

[65] Strawson (2002), p. 452.

grounding true desert is hyperbolic, one will, necessarily, think the same about mine.

Clarke focuses squarely on Strawson's characterization of responsibility "in terms of the desert of punishment or reward".[66] If eternal torment could be deserved, Clarke notes, then it seems that the suffering accompanying it would, by definition, also have to be deserved. So, Clarke understands Strawson as holding "that to believe in 'true moral responsibility' is to be committed to its making sense, at least, to suppose that some of us might deserve such suffering".[67]

In attempting to counter Strawson's argument, Clarke homes in on the phrase "makes sense", which he finds ambiguous. On the one hand, he says, Strawson might be read as merely claiming that the thesis that the guilty deserve to suffer in hell is *intelligible* insofar as one can at least make sense of it, even if it is "patently and necessarily false".[68] But Clarke points out that the same could be said "of the proposition that there might exist something that is both round and square".[69] So, to assert that the guilty deserve to suffer in hell makes sense in this way is not to assert anything particularly interesting.

Clarke accurately interprets Strawson, *not* as arguing that the claim that the guilty to deserve to suffer in hell is intelligible in the minimal sense discussed above but, instead, "that something like [the thesis that the guilty deserve to suffer in hell] might be true is supposed to express the 'scope and force' of true moral responsibility".[70]

---

66 Clarke (2013), p. 10.
67 *Ibid*.
68 *Ibid*.
69 *Ibid*.
70 *Ibid*.

Strawson's claim is that *if* S possessed true moral responsibility, *then* it would at least be *possible* that eternal torment in hell could be a just form of punishment for S. Reasons in virtue of which punishing S in this way would be unjust would have to be *moral* reasons; they could not be reasons relating to the nature of S's *agency*.

Clarke resists the suggestion that "an affirmation of moral responsibility commits one to agreeing that" it is possible that eternal suffering could be deserved, pointing out that there is "a familiar objection that it would be contrary to the nature of divine perfection – and particularly to divine justice – for a perfect creator to subject its creatures to eternal torment".[71] He believes that Strawson must, in order to be justified in claiming that we could make sense of eternal suffering being deserved, provide some argument to motivate this claim. He might, Clarke suggests, defend some version of Jonathan Edwards' defense of hell against the objection that it is contrary to the character of a just God. Edwards' argument goes as follows:

> God being infinitely glorious, or infinitely worthy of our love, honor, and obedience; our obligation to love, honor, and obey him, and so to avoid all sin, is infinitely great. - Further: our obligation to love, honor, and obey God being infinitely great, sin is the violation of infinite obligation, and so is an infinite evil - Once more: sin being an infinite evil, deserves an infinite punishment: such punishment, therefore, is just; which was the thing to be proved.[72]

[71] *Ibid*.
[72] Edwards (1789), p. 4.

The idea would be for Strawson to show that, although something like this argument succeeds, in fact, no one deserves to suffer eternal torment because God does not actually exist. But Clarke holds that "short of showing that this argument from Edwards – or some other argument for the same conclusion – succeeds, Strawson has failed to show that an affirmation of moral responsibility commits one to accepting that the supposition that" the guilty deserve to suffer in hell makes sense in the relevant manner.[73] Clarke argues that the conception of moral responsibility whose impossibility the Basic Argument is meant to demonstrate is not one to which believers in the possibility of moral responsibility are obviously committed.

I do not believe that Strawson needs to vindicate Edwards' argument or any other argument establishing that hell could be a just form of punishment, because his argument is not that eternal damnation *might have been*, but *is not in fact*, a just form of punishment. His argument, rather, is that in order for the story of heaven and hell to make sense – as, to many, it apparently does – there would need to be agents who are ultimately responsible for their actions, which is *impossible*, since nothing could be a *causa sui*. The thought, in other words, is that no one *could* deserve to go to heaven or hell since no one could be ultimately responsible for their actions. (Note that Strawson seems to make a grander claim than I defend here: that, fundamentally, no one deserves anything at all. But I leave open the question of whether the impossibility of ultimate moral responsibility entails the impossibility of anyone deserving anything at all. While I

[73] Clarke (2013), p. 163.

suspect that it does entail that no one could deserve anything in the merit-based sense, I see no reason why it should rule out deserving things in the capacity-based sense.)

I believe that Strawson is on firm ground in holding that only ultimate moral responsibility could ground true desert. For given that we cannot create ourselves, we cannot perform any action that is *ours* in any more substantive a sense than that we can do things that flow directly from our decisions and character. We can of course be *causally* responsible for what we do, but, as we have seen, it seems clear that the concept of *moral* responsibility involves something more than mere causal responsibility. For if moral responsibility were reducible to causal responsibility, we would not expect people to support forms of punishment that are justified on anything other than consequentialist or pragmatic grounds - and we would find it difficult to explain why people tend to think that sanctions cannot be deserved by causal systems that are not adult *homo sapiens*.

It seems to me that the crucial fact upon which to focus is, as I have already said, that praise or blame can be truly deserved in the merit-based sense only if *earned*. But we can only earn something in virtue of what we do. And we cannot earn anything at all, in the deepest sense, if what we do is determined by factors beyond our control in the first place. According to the Basic Argument, all that we do is, in fact, caused by factors ultimately beyond our control. So, Strawson can, I believe, accurately hold that the impossibility of ultimate, or "heaven and hell", responsibility rules out the possibility of true moral responsibility altogether.

Objection 3: The Basic Argument Does Not Undermine Agent-Causal Accounts of Moral Responsibility

It seems that agent-causal libertarian accounts of responsibility come the closest to capturing the variety of responsibility that could ground a guilty person's deserving to suffer. According to this view, the actions taken by fully responsible agents are not to be thought of either as events with causal histories that could, in principle, be traced back to the beginning of the universe, nor as causally indeterminate happenings. Instead, this view holds that full-fledged moral agents are irreducible *substances* in possession of the power to bring about acts without themselves being, in turn, *determined by causes beyond our control* to bring these acts about. The idea would be that, in deciding to go for a run, for example, I, as an agent - not as a mere locus of consciousness - would exercise the unique causal power that I possess, *precisely in virtue of being an agent*, in order to bring about the decision to go for a run. Thus, a proper causal account of my act would not, on this view, attribute my decision either to a random happening or to a chain of causally determined events. Instead, a proper causal account of my act would attribute its occurrence to *me*, as an individual.[74,75] One might think that an agent cause could be identified as the source of her own character and could, therefore, be properly held ultimately morally responsible for her acts. Indeed, Clarke and Pereboom both suspect that agent causalism is not undermined by Strawson's Basic Argument, at least not as

[74] For the original, explicit defense of agent-causalism, see Chisolm in O'Connor (1995).
[75] For a more contemporary defense of agent causalism, see O'Connor (2002).

Strawson states the argument.

As Clarke notes, "agent-causal accounts are sometimes thought to solve the problem of luck that confronts event-causal views".[76] The problem of luck is the problem of accounting for how a merely undetermined event could count as a free action, rather than a chance happening for which no agent could be responsible. By positing actions to be ontologically distinct from events, agent causalists hope "to secure the agent's having a choice about whether her being a certain way mentally is followed by her performing a certain action, even when she so acts because she is that way mentally".[77]

Clarke acknowledges that one might object that agent-causal accounts of moral responsibility fail to solve the problem of luck because, on such accounts, the way that the agent *is* at some time t does not determine how she will *act* at t, and so however she acts is, in the end, a matter of luck. However, even if an agent causal account were modified so as to hold that there *is* a nomological connection between the way an agent is and what an agent does, this would just push the problem of luck back one level. For a proponent of such an account would then need to explain how the way that one is at any given moment could be anything other than a matter of luck. The account would, in other words, have to specify some relevant difference between a manipulated agent, such as René, and the rest of us. Merely being the ultimate, buck-stopping cause of A is not enough to make an agent truly responsible for A. Of course, it is open to the agent-causalists to simply stomp their feet and insist that self-creation is unnecessary for true

[76] Clarke (2005), p. 18.
[77] *Ibid*.

responsibility. But then they would have to concede that a manipulated agent like René can be truly responsible for his actions. The agent causalists would then effectively be *compatibilists* and have their work cut out for them in explaining why we should prefer their metaphysically bloated account of action and agency.

Pereboom defends the coherence of a specific kind of agent causal view, which Strawson refers to as the "Leibnizian libertarian" conception of agent causation.[78] According to Leibnizian libertarianism, reasons (that is, beliefs and desires) only play one part in the causation of a free action. Reasons may incline an agent toward some action, but they do not *by themselves* cause the agent to act. Rather, the Leibnizian libertarian holds that a full explanation of an action must, in addition to the agent's reasons, make reference to the agent's exercising her special power to bring about free actions.[79]

Strawson argues that Leibnizian libertarianism does not in fact provide us with an adequate model of agential freedom. He believes that a fully rational action must admit of a complete reasons-explanation; that such an action must be proximately caused by the agent's reasons alone. On his view, actions that are not fully explicable in terms of reasons cannot truly be up to any agent, because they will be in some respect non-rational. Non-rational actions are, intuitively, not truly up to an agent in any deep sense, given that such actions are, by definition, not performed on the basis of any reason that the agent recognizes. But fully rational actions are, by definition, fully determined by

[78] Pereboom (2001), p. 66.
[79] *Ibid*.

(and thus inevitable in the light of) an agent's reasons (recall that Strawson assumes reasons internalism and therefore counts as rational actions those that flow from an agent's beliefs and desires). So, Strawson believes, fully rational actions are not fully determined by agents but by reasons that agents do not choose.[80] Thus, the familiar Dilemma of Determinism[81] arises:

> (1) Either (a) some action Z was fully determined by the acting agent's reasons, which were in turn determined by causes that preceded the agent's own existence, or (b) Z was *not* fully determined by an agent's reasons.
>
> (2) If (a), then Z was inevitable given the prior state of the universe in conjunction with the laws of nature and, thus, not ultimately up to the agent.
>
> (3) If (b), then Z was in part non-rational, in which case none of the agent's reasons secured Z, and so Z could not have flowed straightforwardly from any rational consideration on the part of the agent.
>
> (4) (a) and (b) are exhaustive of the possibilities.
>
> Therefore,
>
> (5) Z could not have been ultimately up the agent.
>
> Therefore,
>
> (6) No action can be ultimately up to any agent.

[80] *Ibid*.

[81] The Dilemma of Determinism is an old argument for the impossibility of free will. It can be stated simply as follows: (1) Either an act was determined or it was not; (2) If the act was determined, then it was inevitable, and thus not free; (3) If the act was not determined, then it was random, and thus not free; so, (4) No act is free.

Pereboom believes that Strawson would be right to reject the Leibnizian picture as incoherent if it held that the agent's role and her reasons "were wholly independent of each other".[82] But, on Pereboom's understanding of this picture of agent-causal libertarianism, the agent and her reasons are *not* wholly independent in the decision-making process.[83] According to Pereboom, the Leibnizian should be understood as holding that one part of an agent's causal power consists in "the capacity to consider and weigh reasons, and thereby to guide the causing of choices".[84] Thus, he believes, the Leibnizian picture *does* offer a coherent account of how an action can be caused by the agent without being either determined or brought about by causes over which the agent lacks control.

However, Pereboom offers no account of *what goes into* an agent's exercising this special capacity to cause decisions. And the reason for which he might wish to breeze past this detail is apparent: the Dilemma of Determinism awaits him. Suppose that some agent is choosing between two options, A and B. She weighs and considers her reasons, decides that these reasons favor B, and then causes herself to choose B. We then ask *why* she decided that her reasons favored B (assume, for the sake of argument, that it is transparent to the agent why she chose as she did). Her answer will either consist of a series of further reasons (in virtue of which she determined that her reasons favored B), or it will not. If it does *not* consist of a series of further reasons, then she will have to

---

[82] *Ibid*., p. 67

[83] Note that Strawson's case against the agent-causalist would not need to rest upon a denial of this; indeed, nowhere *does he* claim explicitly that agent causalism holds that the role of the agent is separate from the role of her reasons.

[84] *Ibid*.

answer that, ultimately, she chose B *for no reason at all*. But, intuitively, an action performed for no reason at all is not truly free and, therefore, not truly up to an agent.

However, suppose that the agent did decide that her reasons favored B for further reasons *still*. In virtue of what, we may reasonably inquire, was she moved by *those* reasons? Once again, the answer will have to be (a) "in virtue of further reasons still" or (b) "in virtue of no reasons at all". And, again, if (a), then the agent's decision was determined by these further reason - and if (b), then the agent's decision was determined by *no reason at all* (and therefore was not truly up to her).

The Dilemma of Determinism can be pushed back, but not avoided. And positing the existence of some irreducible agent with special causal powers does not make possible any otherwise unavailable form of freedom. It only makes for some needlessly spooky metaphysics.

8. Summary

I have in this chapter endeavored to show that the widely accepted claim

(G) The guilty deserve to suffer

is false if (G) is understood to express that the guilty *truly* deserve to suffer; that is, for it to be intrinsically good that the guilty should suffer. I then argued that *if* it is intrinsically good that the guilty should suffer, then this must mean that it is intrinsically good that the guilty should suffer either because they have *earned* their suffering through some past action or because they have some *capacity* in virtue of which they ought to suffer. I explained why I believe it to be impossible that anyone could truly deserve to suffer - or, indeed, could truly deserve *anything* bad - simply in virtue of possessing some capacity. So, I concluded, for S to truly deserve suffering, S would need to have *earned* S's suffering in virtue of having responsibly performed at least one past action.

I then considered a recent defense of the claim that the guilty deserve to suffer by Randolph Clarke. Clarke, as we saw, argues that suffering might be deserved because it is intrinsically good and fitting that the guilty should suffer by feeling guilty. I argued that both of these arguments are problematic even setting aside the matter of whether we are in fact responsible for our actions in such a manner that we can deserve to suffer for them.

I next argued that merit-based desert requires *ultimate* moral responsibility. Derek

Parfit seems to agree with me on this point, and I presented his argument for the negation of (G) which appeals to the idea that we are not ultimately responsible, and thus cannot deserve to suffer, because our actions are merely events in time. I argued that, although Parfit's argument is sound, it leaves open the possibility that we *could* bear merit-based desert for our actions if our acts were not merely events in time. I argued that this is untrue and that, in fact, the reason we cannot deserve to suffer is that we *could not possibly be* ultimately morally responsible for the way that we are, mentally. I defended this claim by presenting Galen Strawson's Basic Argument. Finally, I defended the Basic Argument against the charges that (a) it is missing a premise, (b) trades on a hyperbolic conception of moral responsibility, and (c) does not undermine agent-causal accounts of moral responsibility.

References

Chisolm, Roderick. 1995. "Agents, Causes, and Events: The Problem of Free Will" in *Agents, Causes and Events: Essays On Indeterminism and Free Will*. Compiled by Timothy O'Connor. New York: Oxford University Press.

Clarke, Randolph. 2005. "On an Argument For the Impossibility of Moral Responsibility". Midwest Studies in Philosophy.

Clarke, Randolph. 2013. “Some theses on desert.*” Philosophical Explorations.*

Dennett, Daniel. 2014. "Reflections on 'Free Will'". Naturalism.org. http://www.naturalism.org/resources/book-reviews/reflections-on-free-will

Diamond, Jared. 2008. "Vengeance is Ours". NewYorker.com. http://www.newyorker.com/magazine/2008/04/21/vengeance-is-ours.

Edwards, Jonathan. 1789. *The eternity of hell torments*. 2nd ed. London: R. Thomson.

Kane, Robert H. 2003. "Free Will: New Directions for an Ancient Problem" in Kane (ed.): *Free Will.* Hoboken: Blackwell.

Kant, Immanuel. 1793/1960. *Religion within the limits of reason alone*. Trans. T.M. Greene and H.H. Hudson. New York: Harper & Row.

Parfit, Derek. 2011. *On What Matters* (Vol. 1). New York: Oxford University Press.

Pereboom, Derk. 2001. *Living Without Free Will*. New York: Cambridge University Press.

Regan, Tom. 1983. *The Case For Animal Rights*. Berkeley: University of California Press.

Spinoza, Baruch. 1675/1985. Ethics, trans. E Curley. Princeton University Press.

Strawson, Galen. 1986/2010. *Freedom and Belief*. New York: Oxford University Press.

Strawson, Galen. 1994. "The impossibility of moral responsibility." *Philosophical Studies* 75, no. 1/2: 5-24.

Strawson, Galen. 2002. "The bounds of freedom." *In The Oxford handbook on free will*, 1st ed., edited by Robert Kane, 441-60. New York: Oxford University Press.

Taylor, Richard. 1974. *Metaphysics*. Engelwood Cliffs NJ: Prentice-Hall.