

University of Richmond

UR Scholarship Repository

Honors Theses

Student Research

4-26-2021

On “The Impossibility of Moral Responsibility”

Dan Jaques

University of Richmond

Follow this and additional works at: <https://scholarship.richmond.edu/honors-theses>



Part of the [Philosophy Commons](#)

Recommended Citation

Jaques, Dan, "On “The Impossibility of Moral Responsibility”" (2021). *Honors Theses*. 1574.
<https://scholarship.richmond.edu/honors-theses/1574>

This Thesis is brought to you for free and open access by the Student Research at UR Scholarship Repository. It has been accepted for inclusion in Honors Theses by an authorized administrator of UR Scholarship Repository. For more information, please contact scholarshiprepository@richmond.edu.

On “The Impossibility of Moral Responsibility”

by

Dan Jaques

Submitted to:

Philosophy Department
University of Richmond
Richmond, VA

April 26, 2021

Advisor: Dr. Nancy Schauber

On “The Impossibility of Moral Responsibility”

I.

In 1994, Galen Strawson published his paper, “The Impossibility of Moral Responsibility”. In this paper, Strawson offers an argument in which he claims to show the impossibility of moral responsibility, regardless of whether or not the thesis of determinism is true; he calls this argument the ‘Basic Argument’ for the impossibility of moral responsibility. A summarized version of the argument runs:

- (1) Nothing can be *causa sui*—i.e., self-caused.
- (2) In order to be morally responsible, one would have to be *causa sui*.
- (3) Therefore, moral responsibility is impossible.¹

The main part of this paper (section III) will be devoted to challenging the second premise of the Basic Argument, and by extension, to challenging its conclusion. On the other hand, I elect to grant the argument’s first premise. To justify this course of action, I simply say that the premise is self-evident. In contrast, the second premise, at a first glance, would seem like an odd, counterintuitive necessary condition for moral responsibility—a condition that virtually no compatibilists, and certainly not many incompatibilists, would endorse; for if it is correct, moral responsibility is *logically* impossible (assuming that self-causation is logically impossible). Indeed, one might argue that moral responsibility constitutes such a fundamental part of the moral and human experience that Strawson’s necessary condition cannot possibly be correct. This line of argument is ultimately the one I will take in challenging the Basic Argument; I will, however, first consider why one might be led to accept the second premise. I will conclude that the reasons which point in favor of accepting the second premise are not

¹ Strawson (1994): 5.

sufficiently persuasive, and that we have just as much, if not more, reason to reject it. It must be noted, too, that Strawson himself does not provide any extensive justification of his necessary condition in “The Impossibility of Moral Responsibility”, so it cannot be claimed that his own reasoning behind the second premise of the Basic Argument would be at all in accordance with my own in laying out the case for it.

Before the main part of the paper, there is a preliminary section (section II). One of my aims in this section of the paper is to situate the Basic Argument within the context of the larger and ongoing debate between compatibilists and incompatibilists. Towards this end, I will be as brief as possible.

Section II will primarily constitute, though, an attempt to get clear as to what Strawson *means* by ‘moral responsibility’ as he uses the term in the Basic Argument and in “The Impossibility of Moral Responsibility” more generally. In other words, I will try to say what Strawson takes the Basic Argument to show the impossibility *of*—for there are numerous things one might mean by the term ‘moral responsibility’. Consider the following passage from “The Impossibility of Moral Responsibility”:

But if one takes seriously the notion of justice that is central to our intellectual and cultural tradition seriously, then the evident consequence of the Basic argument is that there is a fundamental sense in which no punishment or reward is ever ultimately just. It is exactly as just to punish or reward people for their actions as it is to punish or reward them for the (natural) color of their hair or the (natural) shape of their faces.²

Although they are obviously related concepts, it is at first puzzling to think about what the precise relationship between moral responsibility and punishment or reward may be, and just as puzzling is that fact that Strawson takes the impossibility of moral responsibility to *imply* that *any punishment or reward is necessarily unjust*. Thus, taking

² Strawson (1994): 15-16.

the time to first spell out what exactly Strawson means by 'moral responsibility' will both serve to make the connections and relationships between these concepts (as he employs them) more apparent, and, in particular, to give us a better understanding of why Strawson takes it to follow, as a consequence of the Basic Argument, that no punishment or reward can ever be just. We will also then be in a better position to criticize the Basic Argument and its premises in the following part of this paper.

The final part of this paper (section IV) will briefly concern itself with some of Strawson's other, and somewhat unrelated, puzzling remarks regarding the nature of moral responsibility. Consider:

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.³

How could one justly be punished with something like eternal torment in hell, or rewarded with something like eternal bliss in heaven? (But without being morally responsible, one cannot justly be punished at all?) Even if one is *causa sui*, as Strawson holds one must be in order to be morally responsible, how could such seemingly extreme punishments and rewards *ever* be morally justifiable? Of course, I say that these sorts of punishments are *seemingly* extreme because, if Strawson is correct, then the sort of punishments and rewards in question would not be extreme, but rather, *deserved*. But, how could such harsh treatment ever be deserved?

Before turning the body of the paper, I will note that in what follows, I focus exclusively on moral responsibility *for actions*. Additionally, I only focus on the *negative*

³ Strawson (1994): 9-10.

side of moral responsibility for actions, and the attitudes, practices, and actions associated with it. That is, I will limit my focus to the negative reactive attitudes that express censure, such as blame, resentment, indignation, and reprobation; harsh forms of treatment such as punishment; and bad or faulty actions. I limit myself in these two respects primarily for the sake of convenience. However, in principle, (though some may disagree) what is said of these negative aspects of moral responsibility should apply equally well to the corresponding positive aspects—i.e., of praise, approbation, reward, good or virtuous actions.

II.

I now set out to situate Galen Strawson's Basic Argument within the larger context of the ongoing debate between compatibilists and incompatibilists. Once I have done this, I will turn my attention towards the primary aim of section II; namely, to say what Strawson takes the Basic Argument to show the impossibility *of*. Of course, we may now say that Strawson takes the argument to show the impossibility of *moral responsibility*, but we still may be unsure—and at this point we *are* unsure—as to what he means by this. In laying out the historical debate between compatibilists and incompatibilists, though, I will write as if under the assumption that we have a working understanding of what is meant by 'moral responsibility', even though this may not be so.

Historically, throughout the twentieth century, the debate between compatibilists and incompatibilists was primarily centered around what was called the 'could-have-done-otherwise' condition, or the 'principle of alternate possibilities'. It was generally thought that in order to be morally responsible for her action, an agent had to

have alternative courses of action that were *genuinely* open or available to her at the time of her choice to act in a particular way—i.e., that an agent had to *genuinely* have been able to ‘act otherwise’, whatever this might mean. Incompatibilists, on the one hand, thought that the truth of determinism would preclude the possibility of ‘being able to act otherwise’, while compatibilists, on the other, held that it would not.

However, towards the end of twentieth century, incompatibilists began to focus on a *separate* necessary condition of moral responsibility—a condition which they held to be more fundamental than the ‘could-have-done-otherwise’ condition. This shift in focus, away from the ‘could-have-done-otherwise’ condition, occurred largely in response to Harry Frankfurt’s paper, “Alternate Possibilities and Moral Responsibility”.⁴ Frankfurt seemed to demonstrate, through the use of what came to be known as ‘Frankfurt-style examples’, that one could have been genuinely unable to act otherwise while performing an action, and yet still be correctly assessed as morally responsible for it.⁵ We need not get into the details of these examples in this paper—it will suffice to say that many incompatibilists accepted Frankfurt’s arguments, and subsequently, because of this, they turned their attention towards a separate necessary condition of moral responsibility.

In order to be morally responsible for her action, these incompatibilists maintained, an agent must be *ultimately responsible* for bringing it about (henceforth referred to as the ‘U-condition’ of moral responsibility).⁶ This means, approximately, that the agent must be responsible for the ultimate causes of her action—i.e., the determining or non-determining causes at the start of the causal chain which result in

⁴ Frankfurt (1969).

⁵ Boxer (2013): 12.

⁶ I take the term ‘U-condition’ from Boxer (2013).

her action's occurrence. Strawson, falling into this class of incompatibilists who endorse some form of a U-condition, endorses a particularly strict formulation of this condition, according to which no agent can be morally responsible for their actions. That is, he maintains that in order to be ultimately responsible for one's action (and thus *morally responsible* for one's action), one must be *causa sui* (the second premise of the basic argument). On this formulation, the condition is logically impossible to satisfy; and because of this, moral responsibility is logically impossible.

Not all incompatibilists formulate the U-condition in this way—in fact, most do not.⁷ Some incompatibilists endorse something like the initial approximation of the U-condition above. On this approximation, it appears that the U-condition may be satisfied by some sort of agent-causal theory of human action.^{8,9} Some other incompatibilists posit that indeterminism, when inserted at certain key points in the causal chain which brings about an action, may amount to an agent's ultimate responsibility for the action, and this would amount to an agent's moral responsibility for it.¹⁰ Strawson, though, is steadfast in maintaining that an agent must be *causa sui* in order to be morally responsible. He is thus a member of a subclass of incompatibilists who reject the possibility of moral responsibility altogether. In other words, he is a skeptic about moral responsibility.

⁷ For a discussion of the different ways in which incompatibilists have formulated U-conditions, see Boxer (2013): 12-35.

⁸ See O'Connor (2000).

⁹ Roughly, an agent-causal theory of human action maintains that an agent, when she causes an event such as an action, is not brought to act by any antecedent causes. The agent, under such a theory, is thought of as a persisting substance, and as such, does not exist as the effect of some prior cause. One task of the agent-causal theorist is to explain how agents, viewed as 'agent-causes', are able to act on reasons, i.e., act rationally, without allowing what the agent takes as a reason to figure into the *causal* explanation of the agent's decision or intention to act in a certain way.

¹⁰ See Kane (1996).

Compatibilists, on the other hand, reject the idea that the satisfaction of a U-condition is necessary for moral responsibility. Regarding Strawson's position with respect to the U-condition and moral responsibility in particular, the compatibilist would take issue with the idea that one would need to be *causa sui* in order to be morally responsible. Why, they would press, would one need to be self-caused in order to be morally responsible? Thus, if the Basic Argument is to be adequately defended against the compatibilist, the defender of the Basic Argument must be able to meet the challenge of the compatibilist by supplying good reasons to think that one would need to be *causa sui* in order to be morally responsible. It will be my purpose in section III of this paper to consider the adequacy of these reasons.

—

I now turn to attend to the primary aim of this section of the paper: to make out what Strawson means by 'moral responsibility'. I start by suggesting straightaway that Strawson, in the Basic Argument, is concerned to show the impossibility of what Gary Watson has called 'moral accountability'.¹¹ According to Watson, moral accountability is a particular *kind* of moral responsibility, distinct from other sorts. As I understand it,¹² moral accountability is concerned with holding agents accountable for their moral faults through moral sanction. As such, moral accountability is concerned with both ascriptions of *moral fault* and the *desert of moral sanction*. Strawson, I say, by denying the possibility of moral responsibility, is denying that an agent can ever possess a moral fault for which they would deserve to be held accountable *for* through moral sanction. I

¹¹ Watson (1996).

¹² My views about moral accountability are not equivalent with Watson's, however, I do think that we *mean* more or less the same thing by 'moral accountability'. Where I am following Watson, I do my best to note this explicitly.

will now say a bit more about moral fault and the desert of moral sanction—both aspects of moral accountability.

Moral Fault—the word ‘fault’ is used in a variety of ways. Four different uses, or senses, of the word ‘fault’ come to mind that seem worthy of mention here, though there are likely more: the causal sense, the legal sense, the appraisal sense, and what I propose to call the ‘moral’ sense. Moral accountability, as I understand it, is concerned with the causal, appraisal, and ‘moral’ sense of the word ‘fault’ in some way or another. I will now distinguish these three senses of fault—the moral, appraisal, and causal senses—from each other, and from the legal sense, so as to be clear about which senses of ‘fault’ I take moral accountability to be concerned with, and in which way.

First off, when we use the causal sense of ‘fault’, as we do in the statement ‘the tsunami was the earthquakes’ fault, we simply mean to convey causal responsibility of a direct or indirect sort. That is, we just mean that the earthquake *caused*, or played a causal role in bringing about, the tsunami. Moral accountability is concerned with the causal sense of the word ‘fault’ because, simply, one must have caused their action, in a certain specified sense, in order to be held morally accountable for it.

Next, when we use ‘fault’ in the legal sense, we make statements such as ‘he is at fault for the crime’. By this, we mean that some agent has satisfied some set of conditions according to which he is legally at fault for some crime, and that he is deserving of some sort of legal sanction on this basis. Moral accountability is not concerned with this sense of the word ‘fault’, although the legal sense is analogous to what I call the moral sense of fault.

Third, when we use the word ‘fault’ in the appraisal sense, we make statements such as ‘the engine is faulty’. By this, we mean that the engine is in some way *defective*. It will be noticed that the appraisal sense of ‘fault’ is, in some sense, moral; or more precisely, that it may be used (I follow Watson here) in a way that appears to carry with it something of moral significance—for instance, when it is used to describe human behavior or agency, as in the statement ‘he is a faulty agent’. Yet, this sense of ‘fault’ is not related, necessarily, to the final, and what I have been calling the moral sense of ‘fault’. More on this in a moment.

The final sense of ‘fault’, the moral sense, used to designate what I will call a ‘moral fault’, is, as noted, in some respects analogous to the legal sense. It is used when we make statements like ‘he is morally at fault for the murder’. By this, we mean, that an agent has satisfied certain requirements, usually held to be a combination of causal and capacity requirements, according to which he is at morally at fault *for the murder*, and that he is deserving of some *moral sanction*, on this basis (i.e., a moral fault is the desert base).

A few notes are in order here. First, note that we would appropriately employ the appraisal sense of ‘fault’ to describe the behavior of a murderer. That is, we would appropriately describe his behavior as *faulty*. To describe the murderer’s behavior in this way is, and again I follow Watson here, certainly to do something of moral significance. Yet, this fact alone—that his behavior is faulty, and that it can correctly be described as such—does not *itself* imply that he is *at fault, morally speaking*, on the account of this fact, and therefore deserving of moral sanction. For instance, if the murderer is insane, then he would not satisfy the capacity requirement of moral fault (assuming that one

cannot be insane if they are to be morally at fault), and thus he would not be deserving of moral sanction on account of his murder, even though his behavior is still correctly described as *faulty*. Thus, the appraisal and moral senses of 'fault', as I have distinguished them, are indeed distinct, or serve to pick out different, although related, things.

Second, we might distinguish between an agent's *eligibility for* ascriptions of moral fault and an agent's *actual possession* of moral fault. In order to even be *eligible* for ascriptions of moral fault in the first place, one would have to satisfy certain causal and capacity requirements which, once satisfied, qualify the agent as merely eligible for ascriptions of moral fault. That is, the agent must, presumably, tend to stand in a specified causal relation to their actions and have a certain level of competence in order to be eligible for ascriptions of moral fault at all. Yet, an agent who is merely eligible for an ascription of moral fault need not *actually possess* a moral fault. In order to actually possess a moral fault, an agent must first be eligible, *and* must have *acted badly*. (The agent who possesses a moral fault must have stood in a certain causal relation to their action, *their action must have been bad*, and they must have been competent while performing their action.) We might say, in order to make this distinction clearer, that those agents who are eligible for ascriptions of moral fault are *morally responsible agents*—although, importantly, they need not be morally responsible for any action *in particular*. Those agents who actually possess a moral fault in virtue of some action they have performed are morally responsible agents too, but they are furthermore morally responsible for some bad action in particular—and hence, on my account, deserving of moral sanction on this basis.

It is in this way that moral accountability (a kind of moral responsibility) is concerned with both what I have called the ‘appraisal’ and ‘moral’ senses of fault. In terms of necessary and sufficient conditions, satisfaction of certain causal and capacity requirements are *necessary* for an agent to possess a moral fault and to be deserving of moral sanction, but it is only the satisfaction of these requirements, *taken together* with their having performed a bad action, that is *sufficient* for the agent’s desert of moral sanction.

The Desert of Moral Sanction—according to the above analysis of what I have called ‘moral fault’, an agent is deserving of moral sanction if and only if they are morally at fault. To re-emphasise, this means, in other words, an agent is deserving of moral sanction if and only if he has satisfied certain requirements according to which he is first *eligible* for ascriptions of moral fault, and he has *actually acted badly*. What *it is* to be at fault, morally speaking, is just both to have satisfied all of these conditions *and* to be deserving of moral sanction in virtue of having satisfied these conditions. We might say that the agent’s moral fault is what *grounds* the agent’s desert of moral sanction. By subjecting agents to moral sanction because they are morally at fault, I suggest that we are holding them morally accountable, and that this is what moral accountability consists in.

When speaking of moral sanction, I have in mind certain reactive attitudes that express censure, such as blame, resentment, indignation, or reprobation, as well as harsher, more tangible forms of punishment, that are all, in some sense, *deserved responses* or *reactions* to a moral fault. What distinguishes moral sanction from other kinds of sanction, such as legal sanction, are, first and foremost, the grounds on which

they are deserved; although the actual sanctions which are deserved may be the same. We might distinguish between moral grounds of desert from legal grounds of desert. For example, if one were to claim that an agent is deserving of sanction because he has acted wrongly, this would seem to be a distinctively moral claim, based on a distinctively moral ground, different from the legal claim that an agent is deserving of sanction because he has broken a law of the state, which is based on a legal ground, even though the agent may be deserving of ten years in prison on both accounts. On the first account, I maintain that the agent is deserving of *moral* sanction, because the ground on which he is deserving of the sanction is a distinctively moral ground—that is, what I have called a moral fault.

I do not intend, here, to offer a precise account of the nature of moral desert and desert claims. Of course, though, we may grant that moral desert claims have something to do with justice. It also seems fair to grant that if one is *not* deserving of moral sanction, it would be *unjust* to subject them to such treatment. But, it seems false to say, on the other hand, that if an agent is deserving of moral sanction (i.e., deserving of sanction on moral grounds), it follows that it would be, all things considered, *just* to subject them to such treatment.¹³ At the very least, we might say that if an agent is deserving of moral sanction, there is good reason to subject them to it¹⁴—the agent's being morally at fault (the moral ground of desert) presumably provides such a reason.

Having outlined what I take to be involved in moral accountability, it is plain that Strawson is indeed concerned with the possibility of justly holding agents moral accountability in “The Impossibility of Moral Responsibility”; for ultimately, he denies that

¹³ This point is made by Feinberg (1970): 60, and Boxer (2013): 94.

¹⁴ Again, this point is made by Feinberg (1970): 60, and Boxer (2013): 95.

an agent can ever be deserving of moral sanction. In other words, he denies that an agent can be deserving of sanction on distinctively *moral* grounds. Consider, again, this passage from Strawson:

But if one takes seriously the notion of justice that is central to our intellectual and cultural tradition seriously, then the evident consequence of the Basic argument is that there is a fundamental sense in which no punishment or reward is ever ultimately just. It is exactly as just to punish or reward people for their actions as it is to punish or reward them for the (natural) color of their hair or the (natural) shape of their faces.¹⁵

As we can see, Strawson takes the truth of the Basic Argument to establish that one can never be punished—taken as a sort of moral sanction—*justly*. Yet, this would only follow as a consequence of the Basic Argument if, by ‘moral responsibility’, Strawson means moral accountability as I have described it. For on this understanding of moral responsibility, to be morally responsible for a bad action *just is* to be deserving of moral sanction because that bad action is indicative of a moral fault.

Importantly, because it is ascriptions of moral fault that ground claims of desert of moral sanction, we may interpret Strawson to be denying that we can ever actually have such grounds for such desert claims, i.e., that agents can ever *actually be* morally at fault. That is, Strawson denies that an agent can ever possibly have a moral fault, in the sense described above, for which they could deserve to be held morally accountable *for*.

It is also crucial to note that Strawson does not deny this possibility because he thinks agents can never act badly or perform faulty actions; rather, Strawson is denying that agents can ever be *eligible* for ascriptions of moral fault in the first place. And this is

¹⁵ Strawson (1994): 15-16.

due to his U-condition—agent's are not eligible for ascriptions of moral fault *because* agents cannot be *causa sui*.

On Strawson's account, if he is indeed concerned with moral accountability as I have described it, one is *eligible* for ascriptions of moral fault only when one is (a) *causa sui* (causal requirement), and (b) an agent (capacity requirement). Of course, an agent's satisfaction of both of these requirements is not itself sufficient for an ascription of moral fault and the ensuing desert of moral sanction—the agent must also have acted *badly*. Because the causal requirement on Strawson's account is logically impossible to satisfy, eligibility for ascriptions of moral fault, and thus, desert of moral sanction, is impossible. In other words, *moral accountability* is impossible because there are no moral faults that agents deserve to be held accountable for. It will be my contention in part III of this paper that we do not have sufficient reason to accept this strict causal requirement—i.e., that one must be *causa sui* in order to be deserving of moral sanction, which doubles as the second premise of the Basic Argument.

III.

In this section, I challenge the second premise of Basic Argument, as well as its conclusion. The version of the Basic Argument that I will consider proceeds as follows:

- (1) Nothing can be *causa sui*—i.e., self-caused.
- (2) In order to be morally responsible, one would have to be *causa sui*.
- (3) Therefore, moral responsibility is impossible.

I have elected to grant the first premise of the argument. This course of action seems reasonable because, for all intents and purposes, the premise is self-evident. I will challenge the conclusion of the Basic Argument, then, by only disputing the second premise.

I have claimed in section II that by ‘moral responsibility’, Strawson means moral accountability—that is, a certain *kind* of moral responsibility. I outlined moral accountability as follows. I said that moral accountability, as I understand it, is concerned with holding agents accountable for their moral faults through moral sanction. As such, moral accountability is concerned with ascriptions of moral fault and the desert of moral sanction. If one satisfies the right causal and capacity requirements, then one is eligible for ascriptions of moral fault. Importantly, a moral fault grounds an agent’s desert of moral sanction. If one satisfies the right causal and capacity requirements, *and* one has acted badly, then one *is* morally at fault, and therefore is deserving of moral sanction. By subjecting the agent who is morally at fault to moral sanction, treatment which they deserve on the account of their moral fault, we hold them morally accountable for their moral fault.

If I am right to say that, in “The Impossibility of Moral Responsibility”, Strawson is denying that agents can ever be deserving of sanction on moral grounds because agents can never be eligible for ascriptions of moral fault in the first place, then Strawson is denying that there is ever anything that agents can deserve to be held morally accountable *for*. Thus, moral accountability would be impossible.

The second premise of the Basic Argument can and should be interpreted, then, either as saying: *in order to be deserving of moral sanction* (deserving of sanction on moral grounds), *one would have to be causa sui*, or: *in order to be eligible for ascriptions of moral fault, one would have to be causa sui*.

But why, I press, should we think that one would have to be *causa sui* in order to be deserving of moral sanction? What sort of case can be made for this view? I will now

briefly take up the task of trying to provide some motivation for it, although I will ultimately disagree with this reasoning.

To begin, I note that in general, it is oftentimes (although not always) just to subject an agent to moral sanction when they deserve it. Because moral faults ground desert claims about moral sanction, an ascription of moral fault 'carries with it' desert of moral sanction, which will oftentimes be justly employed. In virtue of this fact about ascriptions of moral fault, we should agree that the principles by which ascriptions of moral fault are meted out should conform to some standard of *fairness*. We can ask ourselves, then, 'Assuming that an agent has acted badly and is of a level of reasonable competence, which sort of conditions need to be met by the agent in order to fairly ground claims about what they morally deserve?'. First, we may submit that one must be of a reasonable competency in order to be eligible for an ascription of moral fault. But, having admitted this, we may turn to the causal conditions in particular: 'What sort of causal relation must the agent stand in to their bad action if they are to be deserving of moral sanction because of it?'. As a response to this question, one can make the case that the only fair causal condition we can accept is Strawson's U-condition—that one must be *causa sui* in order to be deserving of moral sanction in virtue of their bad action.

Let us imagine two accounts of the ways in which our actions may be brought about causally. Call these accounts X and Y. If our actions come about in the way described by account X, then our actions are the result of a causal chain (determining or non-determining) which antedates our birth. If they come about in the way described by Y, then our actions are the result of a causal chain (determining or non-determining) in

which our intention (or substitute some similar mental state) to act is the first cause in the causal chain, and our intentions themselves are uncaused. Let us also suppose that X and Y exhaust the *possible* ways in which our actions can be brought about causally.

If X, then our behavior is the result of a causal chain which long antedates our birth. In this case, even if we act for what seem to be our own reasons, what we take to be a reason is part of a causal chain antedating our birth, and as such, can be (causally) explained (at least in part) by circumstances antedating our birth, such as our genetic makeup or how our parents themselves were raised, and (at least in part) by environmental factors, such as how we were raised or otherwise influenced.¹⁶ If an agent has acted badly, and X is true, it is (seemingly) arbitrary to ascribe moral fault to the agent in particular, for their action is just as much explained by antecedent and environmental circumstances as they are by the agent's choice to act badly; the agent's choice, and entire choice-making ability, are *themselves* explained by the very same antecedent and environmental circumstances as the bad action. Why then, would the agent be deserving of moral sanction in this case, any more than, say, the agent's parents? More generally, why would the agent alone deserve to be subjected to moral sanction on the account of their bad action, apart from the rest of the factors that played a role in bringing the action about? If we say something like 'because the agent has chosen to act badly on the basis of his reasons, he alone is deserving of moral sanction for his bad action', we should have to offer an explanation of why an agent's reasons for action are such special causes, worthy of grounding such desert claims. Such an explanation might understandably seem suspect—an agent's reasons are, in most ways, just like any other causes. And in the case that no such explanation can be given,

¹⁶ Thank you to Dr. Schaubert for helping me to formulate this point.

the argument goes, assigning moral fault to the agent in particular, and saying they alone deserve moral sanction on the account of their bad action, would be arbitrary and therefore unfair, for it is unclear why the agent *alone* would deserve this moral sanction, any more so than the other causal factors which brought about the agent's bad action.

If Y, then our intention to act in a particular way (or substitute a similar mental state) is the first cause in the causal chain which results in our behavior, and our intentions are themselves uncaused. Under this account, it is (seemingly) unfair to ascribe moral fault to an agent on the account of their bad action because *nothing explains* (causally) why the agent formed the intention that they acted upon in the first place. If one were to ask *why* the agent formed the intention that they did, or what explains the agent's intentions, there would be no (obvious) explanation. Even if the agent's intentions were bad, we would seem to want to know *why* they acquired bad intentions in the first place, before assigning them moral fault. If no satisfactory explanation of this sort can be given, it is again unfair to ascribe moral fault to agents under account Y, for it is unclear why the agent would be deserving of moral sanction at all, based on the limited knowledge we have of how the agent's intention to act badly came about.

If, on the other hand, agents were *causa sui*, this would presumably constitute a fair basis upon which we could ascribe moral fault and ground moral desert claims; for the agents would be the original cause of their entire mental constitution. Sadly, being *causa sui* is logically impossible, thus it does not constitute a way in which our actions may possibly come about, causally speaking. So, any attempt to fairly ground moral

desert claims in the ways in which actions may possibly come about causally will ultimately be a failure.

I am unsure of the extent to which Strawson's own reasoning behind the second premise of the Basic Argument would accord with the reasoning I have just provided, for Strawson offers no extensive justification of his second premise—at least he does not in “The Impossibility of Moral Responsibility”. It will be noticed, though, that the above treatment of account X was inadequate. “To ascribe moral fault to the agent in particular”, one might object, “even when his behavior was ultimately brought about as the result of a causal chain that long antedates his birth, *is not arbitrary*. An explanation *can* be given as to why an agent's reasons are such special causes, special enough to fairly ground claims about what the agent morally deserves. So long as the agent is caused to act *by his reasons*, he is eligible for ascriptions of moral fault, and deserving of moral sanction when he acts badly.” One can imagine numerous similar objections to the above treatment of account X. “So long as the agent is caused to act by desires with which he identifies, he is eligible for ascriptions of moral fault” or “so long as the agent is caused to act by his second-order desire, he is eligible for ascriptions of moral fault” are two possible others. Yet, I suspect Strawson may reply to these objections as follows. “Yes, on these accounts, ascriptions of moral fault will not be meted out *arbitrarily*, but they will still be meted out *unfairly*. An agent's reasons for acting (to pick one objection) are themselves ultimately brought about by conditions over which the agent had no control. How is it fair, then, to ascribe to them a moral fault when this is the case?”

Fortunately, Strawson does respond to what he perceives as certain objections to the Basic Argument; namely the compatibilist objection, the libertarian objection, and

the agent-causal objection. His responses to these objections may serve to give us a taste of his reasoning behind the second premise of the Basic Argument, as well as provide a good starting point at which we may begin to challenge the Basic Argument itself. For the purposes of this paper, I put aside the agent-causal and libertarian objections to the Basic Argument, and Strawson's responses to them, so that I can focus primarily on the compatibilist objection to the Basic Argument, as well as Strawson's response to it. For it is my contention that Strawson has not adequately addressed the compatibilist objection to the Basic Argument in particular.

Strawson articulates the compatibilist objection to the Basic Argument as follows:

Compatibilists believe that one can be a free and morally responsible agent even if determinism is true. Roughly, they claim, with many variations of detail, that one may correctly be said to be truly responsible for what one does, when one acts, just so long as one is not caused to act by any of a certain set of constraints (kleptomaniac impulses, obsessional neuroses, desires that are experienced as alien, post-hypnotic commands, threats, instances of *force majeure*, and so on). Clearly, this sort of compatibilist responsibility does not require that one should be truly responsible for how one is in any way at all, and so step (2) of the Basic Argument comes out as false. One can have compatibilist responsibility even if the way one is is totally determined by factors entirely outside of one's control.¹⁷

In response to the compatibilist objection, Strawson says:

One does what one does entirely because of the way one is, and one is in no way ultimately responsible for the way one is. So how can one be justly punished for anything one does? Compatibilists have given increasingly refined accounts of the circumstances in which punishment may be said to be appropriate or intrinsically fitting. But they can do nothing against this basic objection.

Many compatibilists have never supposed otherwise. They are happy to admit the point. They observe that the notions of true moral responsibility and justice that are employed in the objection cannot possibly have application to anything real, and suggest that the objection is therefore not worth considering. In response, the proponents of the Basic Argument agree that the notions of true moral responsibility and justice in question cannot have application to anything real; but they make no apologies for considering them. They consider them because they are central to ordinary thought about moral responsibility and justice. So far as most people are concerned, they are the subject, if the subject is moral responsibility and justice.¹⁸

¹⁷ Strawson (1994): 16.

¹⁸ Strawson (1994): 17.

Strawson is correct to note that compatibilists are not troubled by admitting that one can indeed never be ultimately responsible for what one does, or for how one is, since one cannot be *causa sui*. Yet, I submit, Strawson says nothing of substance to address the actual compatibilist objection; for the compatibilist holds, contra Strawson, that one need not be *causa sui* in order to be morally responsible, i.e., in order to be eligible to receive ascriptions of moral fault and to deserve moral sanction. Strawson, in response to the compatibilist, simply insists that this is not the case—that is, he insists that because we are not ultimately responsible for our actions, we can never be deserving of moral sanction on the account of our actions.

Strawson is surely right in his characterization of the compatibilist's position regarding the causal conditions for grounds for desert of moral sanction, in that compatibilists do tend to claim that an agent is deserving of moral sanction in virtue of their bad action just so long as they were not caused to act by any certain set of constraints, such as kleptomaniacal impulses, etc., but he has said nothing to actually draw us away from this view.¹⁹ He has simply pointed out that agents can never be ultimately responsible for their actions. But is not the compatibilist objecting to the view that one would have to be ultimately responsible in order to be deserving of moral sanction? How, then, is Strawson's response to compatibilist objection at all adequate? Moreover, Strawson says that compatibilist responsibility, because it does not require ultimate responsibility, cannot amount to 'true moral responsibility'. But, if by 'true moral responsibility' he means anything other than ultimate responsibility itself, Strawson

¹⁹ In general, it is best to think of compatibilist moral responsibility as following from the absence of excusing conditions, while incompatibilist moral responsibility follows from a more grand 'metaphysical' freedom, specified in various ways. I thank Dr. Schaubert for making this point.

would seem to be begging the question against the compatibilist—for the compatibilist contents that ‘true’ more responsibility *does not* require ultimate responsibility.

Is there something to the ‘basic objection’ Strawson puts forth against the compatibilist account of moral responsibility? “One does what one does entirely because of the way one is, and one is in no way ultimately responsible for the way one is. So how can one be justly punished for anything one does?” This objection to the compatibilist still begs the question, for the compatibilist contends that one need not be ultimately responsible for the way one is in order to be morally responsible for one’s actions, but it also evokes my remarks above concerning the causal conditions under which it is fair to ascribe moral fault to agents, provided that they have acted badly and are competent.

Recall that under account X, our behavior comes about as the result of a causal chain (determining or non-determining) which long antedates our birth. If determinism is true, then this causal chain would be determining. Call this account X*. If Strawson can make the case that the compatibilist’s more relaxed causal requirement for moral fault would allow for ascriptions of moral fault to be meted out unfairly under X*, then he may have a genuine point against the compatibilist.

Strawson might argue as follows: “Under X*, the way one acts is determined by the way one is. The way one is is determined by antecedent conditions predating one’s birth, over which one had no control. Because the way one acts is ultimately determined by conditions over which one had no control, it is fundamentally unfair to ascribe to one a moral fault on the account of a bad action they have performed. In general, it is unfair to ascribe to anyone a moral fault on the account of something which was determined

by conditions over which they had no control. Because the compatibilist causal requirement for ascriptions of moral fault would permit ascriptions of moral fault to agents on the account of their actions which were determined by conditions over which they had no control, the compatibilist causal requirement is unfair. Because ascriptions of moral fault ground agent's desert of moral sanction, they cannot be meted out unfairly. Therefore, the compatibilist's causal requirement for ascriptions of moral fault is wrong."

Yet, the compatibilist should push back on the principle 'it is unfair to ascribe to anyone a moral fault on the account of something which was determined by conditions over which they had no control', for it is not self-evident, as Strawson might presume.

As far as I am concerned, the compatibilist would be entitled to respond to the above argument in the following way: "It is plain that, sometimes, agents *do* deserve to be subject to moral sanction. That this is so is a fundamental fact about living with other people, i.e., about the human experience. Thus, because any given agent's desert of moral sanction is grounded in a moral fault, people must sometimes be morally at fault in virtue of an action that they have performed. At the same time, if determinism is true, it is likely that *every* action must have been ultimately determined by conditions over which the agents who perform them have no control. So, agents must sometimes be morally at fault in virtue of having performed some action that was ultimately determined by conditions over which they had no control. Therefore, the principle that it is unfair to ascribe to anyone a moral fault on the account of something which was determined by

conditions over which they had no control is false. There is nothing, then, wrong with our preferred causal requirements of moral fault."²⁰

Strawson, in my mind, has no counter to this argument. At the very least, he has given no indication that he may have some counterargument available. As far as a compatibilist is concerned, then, there has not yet been presented a good enough reason to accept the second premise of the basic argument, which again in essence claims that in order to be deserving of moral sanction, one would have to be *causa sui*.

It seems to me that many incompatibilists, Strawson included, deny compatibilism because they are reluctant to concede the falsity of the principle that it is unfair to ascribe to anyone a moral fault on the account of something which was determined by conditions over which they had no control, and they see the truth of compatibilism as entailing the falsity of this principle. Admittedly, the principle does have intuitive appeal. On the other hand, compatibilists, following P. F. Strawson, are so sure that moral responsibility, or the desert of moral sanction, is such a fundamental part of our everyday lives, that they could never accept such a principle in conjunction with determinism, for it would seem to preclude the possibility of moral responsibility, or rule out the desert of moral sanction. In this instance, I am more sympathetic to the compatibilist. That people are deserving of this or that moral sanction in virtue of having done something wrong or objectionable is evidently something we feel regularly. Even though agents cannot have control over the ultimate causes of their actions, we commonly feel nonetheless that others often are deserving of moral sanction on the account of some sort of moral fault that they have displayed. Because this fact of

²⁰ Note that to make this argument amounts to a refusal to countenance skepticism about moral responsibility. Thank you to Dr. Schaubert for explicitly making this point to me.

human and moral life is so self-evident, we should at minimum, I submit, have compatibilist sympathies. And while the argument offered in response to Strawson above may not do much to convince Strawson to embrace compatibilism himself, the fact that it can be made shows that the Basic Argument does not establish that moral responsibility is impossible, for we yet have no good reason, or good enough reason, to accept its second premise.

IV.

In this last section, I wish to briefly comment on Strawson's remark concerning the 'heaven and hell' variety of reward and punishment, which I quoted earlier in this paper. Recall the passage:

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.²¹

Recall too that Strawson understands the morally responsible agent to be *causa sui* necessarily. Adopting again the terminology that I claimed to be associated with moral accountability, I understand Strawson to mean in the above passage that if one is (a) *causa sui* (causal requirement) and (b) an agent (capacity requirement), and that one has (c) acted badly, or perhaps even very badly, then one is morally at fault on the account of their bad action, and is therefore deserving of eternal punishment in hell (some other moral sanction of equal magnitude) in virtue of one's moral fault.

Let us consider this remark in light of what I have interpreted Strawson to mean by the second premise of the Basic Argument—that in order to be deserving of moral

²¹ Strawson (1994): 9-10.

sanction *of any kind*, one would have to be *causa sui*. Taken in conjunction with one another, these two views are admittedly puzzling. For according to the conjunction of both views, and provided that one has met the capacity requirement for moral fault and has acted badly, one is supposedly deserving of eternal punishment in hell if they are *causa sui*, but at the same time, one is supposedly not deserving of any kind or any magnitude of moral sanction if they are not *causa sui*.

Of course, it must be remembered that Strawson denies the possibility of one's being *causa sui*. So, he does not think that one can be deserving of eternal punishment in hell in actuality. But still, we can imagine an agent who is *causa sui*. And we can also imagine that this agent has done horrible deeds. We may ask, then, "Do these facts about the agent—that he is both *causa sui* and horrible—obviously show that he is deserving of *eternal punishment in hell*, or morally at fault *to the degree* that eternal punishment in hell is deserved in virtue of it?". I submit that eternal punishment in hell does not seem *obviously deserved* even in this case, nor is it obviously that the agent is morally at fault to such a degree as would warrant eternal damnation.

Yet, there are many desert claims that *do* seem obviously true. For instance, the agent who (a) has not been caused to act by kleptomaniacal impulses, coercion, ignorance, desperation, etc., (b) is fully competent, and (c) has assaulted her sister, is almost *obviously* deserving of *at least* spending the night at the police station, along with a fair amount of resentment. We see claims such as this made, and be greeted with acceptance, regularly. That the former desert claim is not obvious in this way should cause us to view it with suspicion. Perhaps it is simply difficult to grasp and spell

out the ethical implications of one's being *causa sui*. Or perhaps there is actually good reason to view it with suspicion.

Bibliography

Boxer, K. E. *Rethinking Responsibility*. Oxford: Oxford University Press, 2013.

Feinberg, Joel. "Justice and Personal Desert", in his *Doing and Deserving*. Princeton: Princeton University Press, 1970.

Feinberg, Joel. "The Expressive Function of Punishment", in his *Doing and Deserving*. Princeton: Princeton University Press, 1970.

Fischer, John Martin and Ravizza, Mark. "Introduction", in John Martin Fischer and Mark Ravizza (eds.), *Perspectives on Moral Responsibility*. Ithaca, NY: Cornell University Press, 1993.

Frankfurt, Harry G. "Alternate Possibilities and Moral Responsibility", *The Journal of Philosophy*, 66, (1969): 629-639.

Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person", *The Journal of Philosophy*, 68, (1971): 5-20.

Kane, Robert. *The Significance of Free Will*. Oxford: Oxford University Press, 1996.

O'Connor, Timothy. *Persons and Causes: The Metaphysics of Free Will*. Oxford: Oxford University Press, 2000.

Pitcher, George. "Necessitarianism", *The Philosophical Quarterly*, 11, (1961): 201-212.

Scanlon, T. M. "Learning from Psychopaths", in D. Justin Coates and Neal A. Tognazzini (eds.), *Oxford Studies in Agency and Responsibility*. Oxford: Oxford University Press, 2019.

Smith, Angela M. "Control, Responsibility, and Moral Assessment", *Philosophical Studies*, 138, (2008): 367-392.

Strawson, Galen. "The Impossibility of Moral Responsibility", *Philosophical Studies*, 75, (1994): 5-24.

Strawson, Peter. "Freedom and Resentment", *Proceedings of the British Academy*, 48, (1962): 1-25.

Watson, Gary. "Free Agency", *The Journal of Philosophy*, 72, (1975): 205-220.

Watson, Gary. "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme", in John Martin Fischer and Mark Ravizza (eds.), *Perspectives on Moral Responsibility*. Ithaca, NY: Cornell University Press, 1993.

Watson, Gary. "Two Faces of Responsibility", *Philosophical Topics*, 24, (1996): 227-248.