

5-2017

## Determinism and the Role of Moral Responsibility

Justin Edward Edens  
*University of Arkansas, Fayetteville*

Follow this and additional works at: <https://scholarworks.uark.edu/etd>



Part of the [Ethics and Political Philosophy Commons](#)

---

### Citation

Edens, J. E. (2017). Determinism and the Role of Moral Responsibility. *Graduate Theses and Dissertations*  
Retrieved from <https://scholarworks.uark.edu/etd/1987>

This Thesis is brought to you for free and open access by ScholarWorks@UARK. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of ScholarWorks@UARK. For more information, please contact [uarepos@uark.edu](mailto:uarepos@uark.edu).

Determinism and the Role of Moral Responsibility

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Arts in Philosophy

by

Justin Edens  
University of Central Arkansas  
Bachelor of Arts in Philosophy, 2012

May 2017  
University of Arkansas

---

Dr. Warren Herold  
Thesis Director

---

Dr. Eric Funkhouser  
Committee Member

---

Dr. Richard Lee  
Committee Member

## Abstract

In order to solve the apparent incompatibility between moral responsibility and determinism, it is necessary to understand moral responsibility in terms of the function it plays within moral systems, which is highly similar to the role played by laws within judicial systems. By showing that a conception of moral responsibility based upon desert is metaphysically untenable, a function-based conception will be showed to be much more likely. Furthermore, by considering why the desert-based conception has proven so resilient, insight into the moral responsibility/determinism debate may be possible. Lastly, this paper considers whether the problems with this conception can be solved, and if not, what kinds of problems would need to be avoided in considered alternate conceptions.

## Table of Contents

Introduction.....	pg.01
Section 1	
1.1 Could not have done otherwise.....	pg.02
1.2 Responses to the Problem.....	pg.04
1.3 The Problem of Metaphysical Responsibility.....	pg.10
1.4 Law and Morality.....	pg.17
Section 2	
2.1 Intuitions on Moral Responsibility.....	pg.25
2.2 Belief and Desert.....	pg.30
Section 3	
3.1 Defense of the Current Conception/Why We May Not Need Change.....	pg.34
3.2 Why Change May be Difficult or Impossible.....	pg.39
3.3 Potential Problems with Alternate Conceptions.....	pg.45
3.4 Another Problem.....	pg.52
Conclusion.....	pg.56

## Introduction

If we do not have free will, how can we be morally responsible for our actions? When the thesis of determinism is brought up, specifically in regards to ethics, this question is often asked. The thesis of determinism and moral responsibility are often regarded as being entirely incompatible with one other. But not only is it incorrect to say that they are incompatible, a thorough examination of the what moral responsibility is shows that regardless of whether we accept that we have free will, the answer has no effect on whether we should hold individuals morally responsible for their actions. In fact, once we consider what it actually means to have free will, the requirement that we have it in order to be morally responsible will be shown to be untenable. Given this, and a proper definition of the acceptable circumstances to assign moral responsibility, a robust morality is perfectly possible.

The paper will be structured as follows: the first section will be dedicated to exploring the problem of moral responsibility and what answers have already been given, and explaining what function I believe moral responsibility holds within systems of morality and why this function is unaffected by the thesis of determinism, true or false. The second section will show that, despite the desert-based conception of moral responsibility being metaphysically problematic, it plays an important part in the functioning of current moral systems. In addition, this section will show how the problems and usefulness of the desert-based conception may be the source of the moral responsibility/determinism debate. The last section will be about whether the problems with the current system of moral responsibility are ones that can be successfully dealt with, and if not, what challenges and considerations would need to be dealt with in order for that system to change.

More narrowly, the popular definition of moral responsibility's role, that people should be

held morally responsible because they deserve to be, has major problems in the context of determinism because of the requirements for responsibility it upholds. It is necessary to consider how morality functions, what moral responsibility does to help it fulfill those functions, and whether it is justifiable, if determinism is true, to view moral responsibility purely as a tool of morality.

Therefore, the primary purpose of this paper is to show that when morality, in particular folk morality, is properly understood as a system with certain functions and goals, and moral responsibility as a means for that system to fulfill those functions and goals, the perceived problems between moral responsibility and determinism will disappear. Furthermore, I will show what requirements a system of moral responsibility must meet in order to fulfill these functions, and how it is possible to fail to do so.

Note that at various points in this paper I will reference how moral systems make use of desert-based notions of moral responsibility. In these cases, I am not speaking of any particular moral system, but rather making a generalization about the majority of observed moral systems which, when considering how to determine praiseworthiness and blameworthiness, have as the most important factor what the individual in question deserves.

## Section One

### 1.1 Could not have done otherwise

In order to understand why whether the thesis of determinism is true or false should have no bearing on whether we hold individuals morally responsible for their actions, it is important to understand why determinism and moral responsibility are considered to be incompatible in the first place. First, the version of determinism that is being discussed here is the following: that for

any event, the causal history of that event and the relevant laws of nature, taken together, mean that that event could not have occurred otherwise than how it actually occurred. This also follows for events which are the result of human intentionality. By following the principle of sufficient reason, which states that any and all events must have a cause, we are required to accept that not only must human behavior always have a cause, but human motivations, desires, and the like must as well. As these causes of human behavior are outside the control of the individual in question, the individual has no real control over how they behave. These motivations and desires are what causes an individual to act in whatever way they choose, and since the causes of these are outside of the individual's control, so too are all of the individual's choices. This is held to be true regardless of what factors are said to be the reason for any given human action.

Basically, you can do what you want, but you cannot want what you want.

What does this have to do with moral responsibility? The important phrase here is “could have done otherwise”. The argument is that if we accept that determinism is true, for every single instance of an individual making a choice, it is true that it was impossible for that individual to have chosen to do otherwise than what they did choose. And if, for any given choice, it is only actually possible for them to choose one course of action, they cannot, or should not, be held morally responsible for choosing that action. This is known as the Principle of Alternate Possibilities.

There seem to be two primary considerations why it is necessary for an individual to have done otherwise in order to be held morally responsible for their actions. The first is simply a matter of having a tenable morality. What would be the point, after all, of requiring something of an individual if whether or not they will actually follow the rule is already determined by factors outside of their control? The second relates and compares a determined individual to a coerced

individual. Speaking from intuition, it does not seem correct to hold an individual responsible for their actions when there are factors beyond their control that are driving their behavior. These include instances where the individual is physically incapable of doing otherwise, when the consequences are such that we believe any rational person would only be capable of choosing one of the available options, and when the person does not meet the requirements to be considered a rational agent, such as person that has severe mental illness. In all of these cases, it is standard to suspend moral judgment on the individual's actions and not hold them responsible for what they do.

The unifying factor for these cases is that the individuals could not have done otherwise, or that they could not reasonably be expected to. And if we should not hold individuals morally responsible in these situations, why should we hold them responsible if determinism is true, given that it represents just as strong, if not a stronger, constraint upon our ability to freely make choices?

## 1.2 Responses to the Problem

There have been a variety of attempts to solve this problem, with varying degrees of success. One popular line of argument is to claim that it is possible for determinism to be true while maintaining that it is reasonable to ascribe responsibility/that free will is possible, a position known as compatibilism. In these cases, the most intuitive or common definitions of what is necessary for free will and moral responsibility are rarely used, but instead a different conception of free will or moral responsibility that does not conflict with determinism is substituted. This is the course of action I will be taking in explaining what I believe moral responsibility to be.



For another example of how this problem might be dealt with, it is possible to claim that even though morality and the thesis of determinism seem to be, and perhaps actually are incompatible, it is still possible to make use of the idea of morality and moral responsibility, given the benefit it offers to society. Many of the things that morality condemns and encourages, theft, murder, and violence in general on the one side, and charity, honesty, and kindness in general on the other, are the same kinds of things that most societies would wish to prevent and promote, respectively, regardless of their beliefs in any moral system. So, the answer to the problem of responsibility and determinism is that while it may or may not be fair to hold individuals morally responsible for their actions if the thesis of determinism is accepted, we should continue to do so given that morality remains a useful tool for controlling behavior. Of course, for anyone that sees moral responsibility as something more than a tool, this answer will likely not satisfy them.

Before explaining my own position, it is important to consider and explain other possible views and why their flaws make it necessary to discard them and seek other possibilities. Naturally, this means that libertarians need to be discussed. Libertarians maintain that moral responsibility and determinism are incompatible with one another, and that since agents have free will, determinism must be false. For this reason, it is considered a kind of incompatibilism, as it asserts that free will and determinism are incompatible. Robert Kane is one such person that has written a considerable amount of material espousing incompatibilism. His paper “Two Kinds of Incompatibilism” (1989) is about attempting to provide a reasonable account of incompatibilism, particularly with regards to offering a coherent account of what it means to have free will under incompatibilism. The primary problem with libertarianism with offering such an account is the incompatibilism itself. If it is asserted that free will and determinism are

incompatible with one another, and that agents have free will, it must be the case that determinism is false. As Kane puts it, “they must assume that indeterminism or chance is somehow involved in free choice or action, and it is notoriously difficult to reconcile this requirement with the rationality and control demanded by free, responsible agency” (Kane, 1989, pg.219). His goal is to show that it is possible to have a rational and intelligible account of free will without determinism.

According to Kane, there are, as the title says, two different kinds of incompatibilism. The first are the traditional theories of incompatibilism, the AC or Agent Cause theories. The majority of incompatibilist theories are AC theories. Theories of the second kind are much rarer. They are the TI theories, which stands for Teleological Intelligibility. Because TI theories are so rare, and AC theories so prevalent, AC theories tend to be identified as the entirety of the incompatibilist side of the free will issue, and the validity of incompatibilism as a whole is judged on how successful these theories are. Kane considers this a major problem, for if the AC theories are unable to solve the problems of libertarianism accounts of free will, it might be concluded that these problems are unsolvable, and incompatibilism untenable, when there is another formulation of libertarianism that can possibly avoid those problems.

The primary difference between these two kinds of theories is that the former relies upon non-event, also called non-occurrent, notions of agent causation. TI theories have no such notions. To explain more precisely what he means, Kane borrows a description of AC theories from C. D. Broad. Broad defines agent causation as the agency of an individual that has the following properties, with (c) being the most important feature for AC theories, and the one which defines them:

a) The self or agent is the sole cause of its free choices or actions

b) its causation can be exercised in two directions, to choose (or do) and to do otherwise  
c) its causation of a free choice or action is the causation of an occurrence or event by a thing or substance which cannot be explained as the causation of an occurrence or event by other occurrences or events (Kane, 1989, pg.222).

While all AC theories agree with these conditions as being necessary, what it is that makes the agency non-occurrent depends on the particulars of the individual theories. But while condition (c) is the defining part of all AC theories, it is also the part that causes them the most problems. While it is usually easy to describe normal causation in terms of what event caused what, non-occurrent agency defies all attempts at providing an account for how it works. To even attempt to offer such a description seems to inevitably lead to breaking the condition for non-occurrent agency that has been set.

At this point Kane wonders why AC theories are so set in maintaining these very strict conditions as being necessary for free will. The answers he arrives at are the incompatibilist intuitions on human action, which Kane offers two quotes to explain, the first from Aristotle's *Physics* and the second from Chisholm in order to illustrate:

“The stick moves the stone and is moved by the hand, which is again moved by the man; in the man, however, we have reached a mover that is not so in virtue of being moved by something else” (VIII, z56a6-8).

“Each of us, when we act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing - and no one - causes us to cause those events to happen” (Chisholm, 1966 p.32).

Both of these embody the central intuition which motivates incompatibilist thinking. Though this is not to say that an incompatibilist believes that humans are completely free from

being moved by any other factors than their own will. The question is whether, after those kinds of factors have been taken into account, could there be anything left which satisfies the conditions above? To summarize before he moves on, Kane formalizes the incompatibilism requirements for free will with two conditions, the Explanation and Ultimacy conditions.

(i) (The Explanation Condition) A free action for which the agent is ultimately responsible is the product of the agent, i.e. is caused by the agent, in such a way that we can satisfactorily answer the question "Why did this act occur here and now rather than some other?" (whichever occurs) by saying that the agent caused it to occur rather than not, or vice versa, here and now.

(ii) (The Ultimacy Condition) The free action for which the agent is ultimately responsible is such that its occurring rather than not here and now, or vice versa, has as its ultimate or final explanation the fact that it is caused by the agent here and now (Kane, 1989, pg.226) (Broad, 1962).

Next, Kane begins to discuss the problems he sees with trying to seriously maintain those conditions as a requirement for free will. The main culprit is that in order to successfully meet those conditions, a free action can in no way be determined. This is why indeterminism is necessary for incompatibilism. But libertarians do not want indeterminism, because it threatens incompatibilism as a whole, because a random action is no freer than a determined one. Maintaining the Ultimacy Condition means that the Explanation Condition is difficult to fulfill. Kane does not think that there is a way for AC theories to satisfactorily answer this problem, and the rest of his paper is devoted to the TI theories. However, the versions of incompatibilism that I am arguing against are the AC theories, because the arguments these theories give are the philosophical basis for the desert-based conception of moral responsibility. While the TI theories are supposed to provide an incompatibilist theory that deals with the problems of the AC

theories, they do not seem to avoid the most important problems that Kane brings up with the AC theories, and adds additional problems unique to TI theories. However, I would be remiss in the exact way that Kane criticizes if I did not discuss TI theories at least somewhat.

What the TI theories entail is that, if there are cases of micro-indeterminism, and Kane thinks that we have reason to believe that there are, it should be possible for these cases to cause macro-indeterminism. Particularly, he is concerned with these kinds of indeterminate events occurring in the brain, where they would be able to be a factor in decision making processes. It seems that he has some kind of quantum indeterminism in mind. His argument is that, in cases of weakness of will, and some unknown number of relevantly similar cases, where the agent has reason to do both actions, it is possible that it is indeterminate which they will actually do. As long as the agent has reasons to do both, and they believe that their action is ultimately thanks to some reason they have, Kane believes that this is enough to be acting out of free will. This claim, that certain cases of weakness of will are indeterminate, is an empirical one, as he believes it cannot be explained why an agent resists or succumbs to weakness of will purely based upon the laws of nature and past circumstances. Whether the kind of micro-indeterminism his account requires exists, however, is a difficult point to prove either way. And it is unclear to me how the agent is acting based on their reasons if the ultimate decision is indeterminate, regardless of whether the agent thinks so. Even if Kane is correct, it seems that free will in the way he wants it is only possible in a limited number of circumstances. As a solution to the problem of determinism, it is lacking, seemingly in some of the ways the AC theories are.

### 1.3 The Problem of Metaphysical Responsibility

Even though I espouse a compatibilist conception of moral responsibility, I think that it is necessary to eliminate an entire category of compatibilism before moving forward, so here I will do my best to show that any desert-based conception of moral responsibility is inherently flawed. The idea that we assign moral responsibility, through praise and blame, based upon what an individual deserves is the primary motivation for declaring that moral responsibility and determinism are incompatible. By eliminating these conceptions, I will be that much closer to showing that the conception of moral responsibility we should accept does not conflict with determinism.

Galen Strawson makes a very clear argument in his paper “The Impossibility of Moral Responsibility” for what it means to have free will in the strong sense and why he believes it is necessary to have it to be morally responsible for one's actions (1994). Galen Strawson lays out what he calls the “basic argument” in the following way: nothing can be the cause of itself, but to have moral responsibility for one's actions, it is necessary to be the cause of oneself, at least in certain important mental aspects. Therefore, nothing can truly be morally responsible. Galen Strawson believes, and I agree, that any meaningful attempt to grapple with the problem between determinism and morality must in some way address this problem. Galen offers multiple formulations of this argument, the most formal of which I believe is useful for fully understanding his position:

“1. Interested in free action, we are particularly interested in actions that are performed for a reason (as opposed to 'reflex' actions or mindlessly habitual actions).

2. When one acts for a reason, what one does is a function of how one is, mentally speaking. (It is also a function of one's height, one's strength, one's place and time, and so on.

But the mental factors are crucial when moral responsibility is in question.)

3. So if one is to be truly responsible for how one acts, one must be truly responsible for how one is, mentally speaking - at least in certain respects

4. But to be truly responsible for how one is, mentally speaking, in certain respects, one must have brought it about that one is the way one is, mentally speaking, in certain respects. And it is not merely that one must have caused oneself to be the way one is, mentally speaking. One must have consciously and explicitly chosen to be the way one is, mentally speaking, in certain respects, and one must have succeeded in bringing it about that one is that way

5. But one cannot really be said to choose, in a conscious, reasoned, fashion, to be the way one is mentally speaking, in any respect at all, unless one already exists, mentally speaking, already equipped with some principles of choice, 'P1' – preferences, values, pro-attitudes, ideals- in the light of which one chooses how to be

6. But then to be truly responsible, on account of having chosen to be the way one is, mentally speaking, in certain respects, one must be truly responsible for one's having the principles of choice P1 in the light of which one chose how to be

7. But for this to be so one must have chosen P1, in a reasoned, conscious, intentional fashion

8. But for this, i.e. (7), to be so one must already have had some principles of choice P2, in the light of which one chose P1

9. And so on. Here we are setting out on a regress that we cannot stop. True self-determination is impossible because it requires the actual completion of an infinite series of choices of principles of choice

10. So true moral responsibility is impossible, because it requires true self-determination,

as noted in (3)” (6-7).

Galen considers this an unbeatable argument against the possibility of moral responsibility. The only point that Galen believes it is possible to attack the argument from is by denying that we must be self-caused in order to be responsible for our actions, but ultimately believes that all such attempts will fail. To this end, he considers three responses that have been made against the basic argument along this line, and shows how each fails to truly refute the basic argument, at least in his opinion.

The first is the position of compatibilism that was detailed earlier in this paper, where it is held that as long as one acts without any external constraints, then one is responsible for one's actions. While this does not require (3) above to be true, it still does not get the kind of moral responsibility that Galen Strawson wants. According to Galen Strawson, the kind of moral responsibility that he is interested in is the kind that involves sending people to Heaven and Hell. That is, a level of responsibility where it is acceptable to send someone to either Heaven for eternal reward or Hell for eternal torment as fitting response to their actions. Galen Strawson believes that, while compatibilism does offer a kind of responsibility for one's actions, a kind of utilitarian responsibility that is used because it is useful, he does not think it will ever offer moral responsibility. He does not go any deeper into his examination of compatibilism than this. Galen Strawson has a singular notion of what is necessary for moral responsibility, and the only definitions of moral responsibility that he will accept are those that agree with every point he made above. Any other possibilities he rejects out of hand. It is this stubborn refusal to accept any other possible ways of defining moral responsibility that causes me to have issue with Galen Strawson's philosophy.

The second position that Galen Strawson considers receives a more in-depth examination,



what he refers to as libertarian. Rather than asserting that determinism and free will are compatible, they maintain that they are incompatible, but because we do have free will, determinism must be false. Even so, Galen Strawson does not see how this leads to the possibility of moral responsibility. We can be no more responsible for the result of random or undetermined events than we can be for events that were wholly determined. The basic argument holds regardless of whether determinism is true or false: it is necessary, for any individual to have moral responsibility for their actions, that they are self-caused. There is another problem Galen Strawson has with this solution. Given that it requires that determinism is false, and determinism is unfalsifiable, it is impossible to know whether we actually have moral responsibility. What Galen Strawson means is that even if it is accepted that indeterminism allows for moral responsibility, if we cannot know that the universe is indeterminate, we cannot know that we are morally responsible. The claim he makes that determinism is unfalsifiable is a curious one, but I do not consider it particularly worthwhile to evaluate that claim here, especially since Galen Strawson does not offer any arguments for why this is the case, but simply asserts it as true. In addition, Galen Strawson speaks as if the only way it is possible to be an incompatibilist is to believe that there is an element of randomness to human behavior. While indeterminism is often understood this way, he does incompatibilists a disservice by presuming that this is the only possible understanding.

The last position maintains that it is impossible for a person to be responsible for the way they are, but challenges the assertion that this is necessary in order to be morally responsible. It says that, given a certain conception of the human self, it is possible for humans to be free and responsible for their actions even if they had nothing to do with their character, motivations, or circumstances. It appeals to the phenomenology of choice, that we experience ourselves are free

agents capable of choosing in one way or another. Despite this experience, Galen Strawson still believes that it is impossible for us to be responsible in the way this position claims. Galen Strawson explains it like this: the reason people think themselves to be free and morally responsible is because our self is separate from our character, personality, and motivations, or CPM. So, whenever a person makes a decision, the CPM is still the factor that determines whether they choose A or B. But the self, which is independent of the CPM, can judge these decisions, which allows for responsibility. The problem, according to Galen Strawson, is that what the self decides is a product of the way it is, is just the same as when we make decisions based on our CPM. So, again according to Galen Strawson, this does not solve the problem of the basic argument. It merely adds another layer of complexity to the human decision making process. The fact that we still tend to believe in our own moral responsibility despite the basic argument is simply because of how strongly we experience ourselves as agents that make decisions.

Honestly, the problem with Galen Strawson's response to this criticism shows the overall problem with the entire paper. While I strongly agree that the basic argument is something that needs to be considered for anyone that wishes to engage with the free will and determinism debate, what Galen Strawson does basically prevents any possibility of a successful objection. He defines what is necessary to have moral responsibility in such a way that makes it impossible, and then applies that definition to any attempts to claim that we do have it. For Galen Strawson, the only definition of moral responsibility that he will allow is one that requires the individual to be self-caused, and since it is impossible to be self-caused, except perhaps for God, moral responsibility is impossible. Even when the premise that moral responsibility does not require one to be self-caused is questioned, he presumes it when making his refutation. Galen Strawson's

response would certainly be that this is acceptable, because the definition for moral responsibility that he has offered is the only one that makes sense. But obviously, not everyone agrees with this. If Galen Strawson wishes to truly respond to the objections to his arguments, he needs to not only offer the basic argument, but also offer an argument for why the definition of moral responsibility used in the basic argument is the only reasonable one. If he can manage that, he truly will have an argument for the impossibility for moral responsibility.

Unfortunately, I do not think he can do so. In fact, I believe that with his arguments, he has finally killed the conception he wished to defend once and for all. The picture of morality that Galen Strawson paints leaves us with a dilemma, where no matter what option we take, we are left with troublesome results. It is my expectation that it is possible to be either realist or anti-realist, that is objective or subjective, with regards to morality, and I believe that if we accept Galen Strawson's conception of what is necessary for moral responsibility, we will have problems no matter which we believe is true. Galen Strawson brings up concepts such as "justice" when making his arguments against moral responsibility, saying that it would not be "just" to hold agents responsible for their choices when the reason they made those choices is the result of influences that they have no control over. However, if we accept that it is not "just" to hold anyone responsible for their actions, in order to avoid acting immorally it would seem necessary to have this particular constraint be the only concern of morality. This kind of morality would not allow for any other constraints, given that we could not justly hold any individual responsible for their actions. A morality consisting only of this requirement seems highly counter-intuitive. Surely there is a very widespread intuition that there are at least some instances where it is reasonable to hold a person morally responsible for their actions. So as a realist account, Galen Strawson's view appears to fail to catch our widely-held intuitions.

If we hold that morality is constructed, that is a more anti-realist view, Galen's definition of moral responsibility makes even less sense. Why would we, as the originators of morality and definers of moral responsibility, allow for a definition of moral responsibility that prevents it from ever being applied? Moral responsibility has some level of utility if constructed in a certain way, so Galen Strawson certainly cannot argue that this is the most reasonable conception of moral responsibility available for us to construct.

Neither of these points is intended to prove that Galen Strawson is incorrect for most of the assertions he makes. I simply wish to show that the second move Galen Strawson makes, that of asserting that because metaphysical moral responsibility is impossible, moral responsibility of any flavor is also impossible, is unreasonable. So regardless of whether we believe that morality is “real” in some sense or whether it is constructed, in practice the definition of moral responsibility that requires an individual to be self-caused to be responsible for their actions leads to unreasonable and untenable results. Galen Strawson has offered an excellent argument for why the desert-based notion of moral responsibility is hopelessly flawed, providing me with an opening to offer alternatives.

Before I move on, I think that it is useful to take a moment to explain why I believe that any desert-based model of moral responsibility is affected by Galen Strawson's arguments, if this is not already clear. First, when I say ‘desert-based’ conception of moral responsibility, what I mean is that when an agent commits an act of moral character, either moral or immoral, a desert-based conception would say that the reason and justification for our praising or blaming the agent in question is that there is some metaphysical praiseworthiness or blameworthiness attached to the agent's actions, and therefore to the agent themselves. What Galen Strawson has done is argue, successfully, I believe, that such a metaphysical praise- or blameworthiness

requires of the agent something which is impossible, that they are self-caused. This is the reason for my belief that the success of Galen Strawson's arguments means the elimination of all desert-based conceptions of moral responsibility.

#### 1.4 Law and Morality

Now that I have narrowed down what an acceptable conception of moral responsibility might look like, or rather what it cannot look like, it is time to explain what I believe moral responsibility is and why determinism does not pose a problem to this conception. Both of these, the explanation of my own viewpoint on moral responsibility and why determinism does not affect it, will be the concern for the following section.

First, it is necessary to provide an examination and comparison of the purpose and function of both moral systems and judicial law. By showing that moral systems are relevantly similar to judicial laws in terms of function, and that judicial laws are unaffected regardless of whether the thesis of determinism is true, I hope to show that any tension seen between morality and determinism is based upon false assumptions, mainly regarding what it means to have free will and what is necessary for moral responsibility. That is, based upon a desert-based conception of moral responsibility.

In order to show that determinism does not matter with regards to judicial laws, it is important to examine what they are, and what their purpose is. Though for a moral realist moral facts are, in a way, similar to the natural laws in that both represent, or are supposed to represent, inherent facts about the universe, they are much more similar to judicial laws in practice. So, when I use the word "law", understand that it is in the judicial sense. The Merriam-Webster dictionary defines a law as a "*binding custom or practice of a community: a rule of conduct or*

*action prescribed or formally recognized as binding or enforced by a controlling authority*”, and this is the kind of law that we are interested in.

The important word here is “binding”; laws are meant to act as constraints upon behavior. They either require that certain actions be done, or that they not be done. So, a law against theft is followed if individuals that are under that law do not engage in theft. But this is not the only aspect of laws that we should be interested in. There is also the purpose of a given law, what kind of societal need it is intended to fulfill. In the case of the law against theft, it is intended to be one of the laws that prevent property rights from being violated. The last aspect of laws that we are concerned with is their enforcement, how it is that laws are upheld, both preemptively and after they have been broken. This varies from law to law, but the general means of getting a law to be followed is to have some kind of penalty attached to breaking it. There are, and have been, various penalties for breaking laws, but the aspect they all share is that they are undesirable, and serve to dissuade and penalize law-breaking behavior.

So, how does this relate to the thesis of determinism? At first glance, the “could not do otherwise” problem seems to make laws meaningless. Imagine that we have an individual that is capable both of breaking the law and following, A or not-A. If we assume that the individual in question is physically capable of doing both A or not-A, then the question of which they actually do is up to what they choose. But according to determinism, which choice they make is out of their hands. If they choose to do A, it was not possible for them to have chosen not-A. Likewise, if they actually choose not-A, then they could not have chosen A. If this is the case, it is out of any individual's hands whether they choose to follow the law or not, and it does not matter whether we make laws against stealing or not. Individuals either will or will not steal.

Hopefully, it is obvious what the problem with using determinism as an excuse to not

make any laws is. It is true that given the thesis of determinism, it is impossible for someone to make a choice other than the one they made, given the same available choices and the same situation. They have no free will in the matter; they will always make the exact same choice. But the key phrase here is “given the same choices and situations”. If the situation changes, it is perfectly possible that the individual that always choose to steal before will now always choose not to steal. The difference can be something as simple as there now being a law against theft, and/or a penalty attached to being caught doing so. As the law is only concerned with constraining behavior, there is no need to be concerned with anything else. All that is necessary is for the law and its penalties to be construed in such a way that most individuals will choose not to steal rather than to steal. Given this, the law will be able to fulfill the purpose it was designed to fulfill, as will any other rule or law that one cares to dream up, as long as there is a formulation that will cause an individual's choice to be affected. Whether the thesis of determinism is true or false does not matter. And I believe the same reasoning can be applied to rules of a moral nature. That is, to moral systems.

The popular conception of the role that moral responsibility plays in society is that generally, we give it on the basis of desert: we offer praise because the individual in question deserves to be praised, and blame for the same reason. But given the tension between this conception of moral responsibility and problems with free will, this admittedly natural and wide-spread conception may not be accurate or tenable. In fact, as I believe Galen Strawson shows above, conceptions of moral responsibility of this nature are likely hopelessly flawed. Instead, what if moral responsibility is conceived as a tool of creating and maintaining moral systems? Despite the desert-based notion of moral responsibility being the most popular, intuitive, and wide-spread one, I believe that this popularity is itself simply part of the framework that sustains

moral systems. I will explain this in further detail later.

Once again, it is important to define what I mean by “moral rule” or “moral fact” before moving forward. Unfortunately, these concepts are much vaguer than those of a judicial law, and providing a description of them will require using some terms that do not have a concrete meaning. A moral fact is a fact about an action that that action is either good or bad, at least, when considered under the umbrella of a particular moral system. So, a good action is good, regardless of whether it is understood, accepted, or known. As to the meaning of “good” and “bad”, which can be substituted with terms such as “moral” and “immoral” and so on, while this does not capture the entire meaning of these terms, something along the lines of “more good actions are actions that should always, morally, be chosen over less good possibilities, and more bad actions are actions that should never be chosen, morally, over less bad possibilities” will suffice for my purposes here. What the content of the moral facts is, and whether they even exist, is a matter for someone else. For my purposes, I will simply be assuming that it is possible for moral facts, and that if they do exist, they have some non-vacuous content. What this content is, that is, what actions are in particular good or bad, is irrelevant. It is also irrelevant whether moral facts are understood to be realist or not. The function of moral systems remains roughly the same regardless of how the rules of a given system come to be or the metaphysics of the system.

So, acting in accordance with moral facts or in accordance with moral systems in general, means performing actions that are good and not performing actions that are bad. So far, while the language being used is quite different, the general result seems to be the same as in the case of judicial laws: morality is concerned with constraining behavior. But this is not all that morality is concerned with. To show why, consider the following scenarios: in the first case, there is an individual that never steals, not even a single time. But this is not because they have any



particular beliefs that prevent them from stealing, it is simply that they never have had an opportunity to steal anything, and never will. However, if they did have such an opportunity, they would, with certainty, choose to steal. In the second case, the individual in question also never engages in theft. In their case, however, it is not because they never have an opportunity to do so. In fact, they have a plethora of chances to steal, but choose never to do so because of the beliefs they hold on the morality of stealing. In both cases, the individual in question never steals anything. If all morality was concerned with was constraining behavior, these cases should be basically identical. But I have the intuition, as I suspect the majority of others do, that the individual in the second scenario is acting morally, while the one in the first scenario is not.

Because of this, I believe that morality concerns itself with not just constraints upon the behavior, but also upon the will. It is not enough to do good things; they must be done for good reasons. So, the question becomes, how do moral rules get enforced? In the case of judicial laws, the government the individual lives under fulfills the role of making sure that they make choices that follow the laws, or by at least offering penalties for breaking those laws. Is there anything similar that enforces moral rules? Unless the state concerns itself with matters of morality, which it does on occasion, no, there is no formal power enforcing it. This is where holding an individual morally responsible, through use of praise and blame, for their actions does work. In any social network where a particular set of moral rules are held as being true, in a sense, every member that upholds those rules, and expects them to be upheld in others is an enforcer of those rules, and holding others morally responsible is the method that is used through the actions of praise and blame. Whenever a member of the system engages in moral behavior, that is, they make a choice to do something that is recognized as either being “good” or “bad”, they receive a response from other members of the group. If the action is “good” then they are praised and told

that whatever they choose was correct. If they choose a “bad” action, they are scolded and blamed, and told not to perform the behavior again. By doing this, members of the group are caused to not only avoid doing “bad” things and to actively pursue doing “good” ones, they also come to believe internally that they should do so. Given that to act in accordance with a moral fact simply is to act morally and will morally, both of these constraints can be properly fulfilled by holding individuals morally responsible for their actions. The further use holding others responsible in this way is that by holding them responsible for their actions, they will often come to believe that the given actions really are good, and/or really are bad, which fulfills the will requirement of acting morally.

Why does praise and blame lead to having a “proper” will? The process, as I imagine it, works like this: when a person is exposed to a particular truth, and there is a high degree of consensus regarding that truth and little dissent, it is only natural for that person to accept that truth themselves. Furthermore, the act of blaming or praising makes the belief that others believe this way all the stronger. So, if a person is told that a particular action is moral or immoral, receives little to no contradictory information, and is lead to believe through the actions of the others that their belief is genuine, they will likely form the same belief. There will likely be little in the way of contradictory information if the truth being propagated is that of the dominate morality.

Of course, this is not guaranteed. There will be times when, for various reasons, the fact that everyone else at least seems to believe a certain way is not enough for a person to believe themselves. But for the most part, it works well enough that the conditions for following the morality, both in action and in thought, are met.

Now, let us look to the thesis of determinism and see if, answered one way or another,

there will be an effect on the ability of individuals to act and will morally. Once again, determinism holds that when a person is given a set of choices, the choice they will ultimately make is determined by factors they cannot control. As was the case when discussing judicial law, this seems to offer superficial reason as to why morality would falter if we hold that determinism is true. An individual lacks the free will to choose to do good things and choose not to do bad things. This, however, is no more a problem here than it was in before: if we wish for someone to choose to do moral actions, we should use the tool of moral responsibility, through praise and blame, in order to guide them into doing so. In this way, they will choose to do what is moral in the future, and choose to do so for moral reasons, as the result of a moral will.

This is, I believe, all that moral responsibility is. A means to cause individuals, through praise and blame, to have certain beliefs regarding a set of moral acts, and to behave in accordance with these beliefs. It is the way that any moral system functions, regardless of the particular tenets that system holds as true. The desert-based conception is one which is completely false, though there is still, I think, good reason that our belief in it has remained so stubborn. Now, while I consider moral responsibility to purely be a tool of moral systems, as to what morality itself is and what its purpose is I am agnostic. Perhaps morality itself is a tool of society used to encourage some behaviors and discourage others, or perhaps it is something that we must follow in order to meet some kind of primitive criteria for 'goodness'. What role moral responsibility has is the same regardless.

It is important to note here that I still think that metaphysical responsibility is separate from moral responsibility, and that metaphysical responsibility is impossible just as Galen Strawson claimed it is. The difference is, is that I do not believe that these two kinds of responsibility should be considered identical. This opens up the problem of justifying holding

individuals morally responsible when they cannot be metaphysically responsible, which I will discuss in more depth in section 3.1.

Reading this description of moral responsibility, an immediate objection that might be had is that according to this theory, it is perfectly acceptable, under certain circumstances, to praise or blame someone for something that they did not do. To this, I have two things to say immediately, and a third to say much later. First, as a description of how moral responsibility actually works, I believe that this result is true to reality. As a general rule, if you wish to encourage some behaviors and mindsets and discourage others, you do not wish to reward or punish those that have not committed the relevant actions. However, there are going to be cases where overall, the amount of adherence to a system will be increased by doing this. In order to satisfy the public's need for clear heroes and villains, society on occasion does this. And if the result is a net benefit, this theory of moral responsibility says that this is acceptable. This is mitigated by the fact that, if you punish or reward against the rules of the moral system you are attempting to propagate, the result will obviously be counterproductive. This is especially the case given that true belief in the system's rules is one of the results being sought. So, this kind of behavior would, by necessity, be rare.

That this might be considered a problem in the first place brings me to my second point: praising and blaming someone for something that they did not do is, according to this theory, no better and no worse than praising or blaming some for something they did do. The entire reason I have offered this theory of moral responsibility is because I think that Galen Strawson is correct when he says that the conditions for us to be metaphysically responsible for our actions is impossible to meet. So, anyone that raises this objection may be, in a sense, missing the point, and it seems likely that they are still ascribing to the idea that it is acceptable to praise or blame

someone for something they did do, metaphysically. No one deserves praise or blame, regardless of whether they committed a particular action or not. To anyone that remains unconvinced, I will also be returning the matter of justifying this kind of behavior in section 3.1.

I think that, given that I am talking about what I consider to be folk or the common morality, it might be useful at this point to bring in some actual empirical data on pre-philosophical beliefs regarding moral responsibility and determinism, in order to show why I think the debate between the two has lasted so long. This should also be helpful in showing, more precisely, how I believe current moral systems function in terms of how agents view moral responsibility and its role in morality.

## Section 2

### 2.1 Intuitions on Moral Responsibility

It is time to engage in the useful, though perhaps not absolutely necessary, task of considering why conceptions of the free will and moral responsibility problem as Galen Strawson has outlined are so popular. That is, why there are intuitions that pull us in different directions. If we accept certain truths about our universe including the principle of sufficient reason, which I suspect the vast majority of all persons, let alone philosophers, do accept, then we are pushed in the direction of accepting determinism. Yet I also suspect that the majority has strong intuitions on the matter of holding individuals morally responsible for their actions, and while they may differ somewhat person to person, it is likely that most will agree that there are circumstances in which morally responsibility is appropriate. Discovering what intuitions non-philosophers have should help in answering why this problem occurs. I am also interested in the intuitions of the average person given that this account is meant to be an explanation of why folk

morality does not conflict with determinism, and whether a favorable interpretation is possible. This is primarily an empirical project, and luckily for me someone else has already put in the effort to gather the relevant information. In “Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility”, (2005). E. Nahmias et al take it upon themselves to see what exactly the general public's thoughts are on the matters of free will, determinism, and “could not have done otherwise”.

The first study, in section 2.1, Nahmias et al performed was what they referred to as the 'Jeremy Cases'. The basic premise of the thought experiment offered goes like this:

“Imagine that in the next century we discover all the laws of nature, and we build a supercomputer which can deduce from these laws of nature and from the current state of everything in the world exactly what will be happening in the world at any future time. It can look at everything about the way the world is and predict everything about how it will be with 100% accuracy. Suppose that such a supercomputer existed, and it looks at the state of the universe at a certain time on March 25, 2150 AD, 20 years before Jeremy Hall is born. The computer then deduces from this information and the laws of nature that Jeremy will definitely rob Fidelity Bank at 6:00 pm on January 26, 2195. As always, the supercomputer’s prediction is correct; Jeremy robs Fidelity Bank at 6:00 pm on January 26, 2195” (566).

First, participants were asked whether they believed such a scenario was even possible. The majority responded they did not think it was. After being told to suspend belief, it was asked whether they believed Jeremy was acting out of his own free will. 76% believed that he was. In order to determine whether the moral aspect of the action was affecting the outcome, saving a child and going jogging were also used, as a positive and neutral action. In both cases, the majority responded that Jeremy was acting in his own free will, 68% for saving the child and 79% for going jogging. Then they turned to blameworthiness and praiseworthiness. 83% considered him blameworthy in the negative case, and 88% considered him praiseworthy in the positive case. Next, participants were tested on whether they believed Jeremy had the ability to do otherwise. For the negative case, 67% responded that they believe he could have done

otherwise, in line with the responses received on free will and responsibility. However, for the positive and neutral actions, the results were quite different. 62% believed he could not help but save the child, and 57% believed he could not help but go jogging. The interpretation that Nahmias et al offered is that the participants had intuitions that agree with authors such as Frankfurt, who will be discussed later, that it is possible to be responsible for one's actions even if one cannot do otherwise, an interpretation that I agree with.

A second scenario was also constructed after the possibility that the first did not make the deterministic nature of the problem salient. The second scenario, found in section 2.2, is focused on presenting the problem in such a way that it should be obvious that the individuals are controlled by factors outside of their control:

“Imagine there is a world where the beliefs and values of every person are caused completely by the combination of one’s genes and one’s environment. For instance, one day in this world, two identical twins, named Fred and Barney, are born to a mother who puts them up for adoption. Fred is adopted by the Jerksons and Barney is adopted by the Kindersons. In Fred’s case, his genes and his upbringing by the selfish Jerkson family have caused him to value money above all else and to believe it is OK to acquire money however you can. In Barney’s case, his (identical) genes and his upbringing by the kindly Kinderson family have caused him to value honesty above all else and to believe one should always respect others’ property. Both Fred and Barney are intelligent individuals who are capable of deliberating about what they do.

One day Fred and Barney each happen to find a wallet containing \$1000 and the identification of the owner (neither man knows the owner). Each man is sure there is nobody else around. After deliberation, Fred Jerkson, because of his beliefs and values, keeps the money. After deliberation, Barney Kinderson, because of his beliefs and values, returns the wallet to its owner.

Given that, in this world, one’s genes and environment completely cause one’s beliefs and values, it is true that if Fred had been adopted by the Kindersons, he would have had the beliefs and values that would have caused him to return the wallet; and if Barney had been adopted by the Jerksons, he would have had the beliefs and values that would have caused him to keep the wallet” (570).

Interestingly, the results of this study are very similar to the results of the previous one.

76% answered that they believed Fred and Barney acted out of free will, and 94% answered the same as they did when asked whether the action was taken freely: 60% judged Fred blameworthy

and 64% judged Barney praiseworthy. Lastly, when asked whether they could do otherwise, 76% answered that they could.

Nahmias et al considered these results to strongly suggest that pre-theoretical individuals do not have incompatibilist intuitions. However, there are some points to consider when interpreting this information. First, simply because they do not seem to have incompatibilist intuitions does not mean that they support compatibilism. Second, the studies were restricted to college students for participants, and the results could potentially be much different should the research be conducted again with a more diverse backgrounds and statuses. It may also be possible that the determinism in these scenarios was not salient enough to the participants. Many participants in the Jeremy case believed that the scenario in question is impossible, which may have resulted in miss-firing intuitions. It is also possible that individuals are so attached to the concept of having free will and being morally responsible for one's actions that they are unable or unwilling to reason that determinism may be incompatible with both or either.

If I am being perfectly honest, before I read the results of these experiments, I was similar to Nahmias et al in that I expected the majority of individuals to have incompatibilist pre-theoretical intuitions. In fact, this seems to be a very popular position to hold, which makes this kind of research all the more important for determining what exactly the intuitions of the 'average person' are. Rather than discover the root of incompatibilist intuitions, I have found that the majority of individuals, at least in these studies, do not have such intuitions at all. While it is possible, as Nahmias et al have said, that the determinism in the scenarios was not salient enough for the participants, it is also possible that it was perfectly salient, at least in the second scenario. For the first scenario, it may be that the participants are not considering the machine to be particularly constraining as it is certainly possible the machine is simply getting lucky. Again, the



principle of sufficient reason is something that I expect the vast majority of individuals to accept, and once it is known and understood it requires very little effort to apply to persons as well. Yet despite this, we continue to utilize various forms of blame in our interactions with each other. It is therefore only natural that pre-theoretical intuitions not be incompatibilist. That is, the kind of moral responsibility that the participants were working off of may already have contained within it the kind of determinism Nahmias et al has in mind. Even if it were only subconscious for the most part, it should not be surprising if individuals have a definition of moral responsibility already accounts for the kinds of scenarios presented here. It is likely that Nahmia et al believed that the determinism was not salient enough because of the responses given when asked whether the individuals in question could do otherwise. Given how the scenario is structured, with an all-predicting machine in the first and a completely determined world in the second, how could the participants claim that the individuals could have done otherwise? I think it is because they are able to operate using a different kind of 'could have done otherwise' that the philosophers are assuming. It is likely that, if the 'could have done otherwise' question was phrased as, "if the personality of the individual, their environment, genes, etc., and their circumstances were exactly the same, would they make the same decision?". To this question, I think they would answer that they would, given the previous responses. The response they may actually be giving is that given a change in circumstances or a change in the personality, the individual could have acted differently than they did, which seems to be in line with the kinds of responses given. Still, it seems that from the percentages of responses going each way, there is nothing approaching consensus on these matters.

But as Nahmias et al stated more research, even significantly more, would be necessary to provide anywhere near conclusive results, for as it stands the data is open to various

interpretations. All that I have offered so far is one such interpretation. Regardless, I think this study provides an insight into what definitions of moral responsibility society operates with, and how people believe that it is correctly and incorrectly assigned.

## 2.2 Belief and Desert

Now that I have provided a bit of empirical data on pre-philosophical intuitions regarding moral responsibility and how I believe that that data should be understood, now seems to be a good time to explain what exactly I think may be the primary motivation for the free will/determinism conflict in further detail, and why I also think that it may be difficult to avoid this conflict. Once again, in order for an action to be moral, it must be done with a moral will. So, an individual that does something moral must, as a strict requirement, have certain beliefs and motivations when they commit that action in order for it to be moral. Part of those beliefs is that they are certain that some actions are moral and others immoral. The individual's beliefs regarding the metaphysical state of morality does not matter. So, whenever these individuals see actions that they consider to be moral or immoral, there is a natural pull to believe that the individual who committed these actions are deserving of praise if the action is moral or blame if the action is immoral. This is a notion of moral responsibility based upon what individuals deserve, and is the same kind of moral responsibility that Galen Strawson espouses. The question I had when reading Galen Strawson is why he seemed so intent on maintaining that particular definition as the only possible one for moral responsibility. While I agree with him that that definition needs to be considered why entering the determinism and free will debate, he does an excellent job of showing why that definition is one that is impossible to fulfill. Instead of clinging to an untenable definition, why not consider other possibilities? I think that P. F.

Strawson, who will be discussed in the next section, does an excellent job of explaining how entrenched the notion of a desert-based system of moral responsibility is, and so for the majority of persons, it is no longer possible for them to seriously consider any alternative notions, or it is at least very difficult. Moral responsibility, and desert-based notions of responsibility in general, are very important tools for moral systems, and in order for these tools to work as they need to, the individuals ascribing responsibility must believe that that responsibility is deserved, for a moral will requires such belief. This is how ascribing responsibility is justified both to the self and to others part of the same system.

But people are capable of reason as well, and that capacity for reason will inevitably cause people to examine any systems they are part of, and such an evaluation on a desert-based system of moral responsibility will likely lead to the same conclusions that Galen Strawson had; that it requires of us things that are impossible. We consider ourselves and the world, and realize that humans cannot be responsible for their actions in the way that is required by desert-based responsibility. So, we are pulled in two directions: by our morality, which we wish to maintain because it is part of how our society functions, we are pulled to believe that it is only acceptable to ascribe moral responsibility when the individual in question is metaphysically responsible what their actions, and by our reason, that tells us that to be responsible in such a way is impossible. This, I think, is the basis for the free will and moral responsibility debate. I do not think this problem will go away, either. As P. F. Strawson says, our current notion of the requirements for moral responsibility are a very important part of our current society, and to change them would require a massive paradigm shift. And so long as people are able to think, they will not fail to notice, consciously and unconsciously, how a person cannot be responsible in such a way.

While fretting over the free will and determinism debate is generally considered the domain of philosophers, and not of particular concern to others, the previous research on non-philosophical intuitions on the subject point in a different direction. For the first experiment, the majority held that even if all of an individual's future behavior is being predicted correctly, and that individual cannot help but do what they did do, they can still be held responsible for their actions. Likewise, for the second scenario, which incorporated what might be considered the factors of classical determinism, such as genes and upbringing. What is important about these studies is the number of individuals that went against the majority, who believed they are responsible, when it was asked whether the person in question is responsible for their actions in the second case: roughly 40%, which is definitely not insignificant. Also interesting is how it departs from the percentages in the first case, in which roughly 85% considered the person in question blameworthy.

If I may be permitted, I will offer a possible answer for this difference that goes with what I have said previously in this section. For the first scenario, the fact that the machine is predicting behavior, even correctly, is being disregarded. This is supported by the responses of the participants, the majority of which did not believe such a machine is even possible. Because the 'determinism' aspect of the scenario is not salient, as Nahmias et al feared, the answers the participants give are based on the normal desert-based notion of what is required for moral responsibility. Because of this, they answer to hold the individual in question responsible. In the second scenario, the way determinism is offered is the way that applies the Principle of Sufficient reason to human action, making it salient. Now that the determinism is salient, roughly 25% more participants respond that the persons in question should not be held responsible for their actions. The question is, why did the participants' intuitions change when the determinism

of the scenarios is made more salient? One possible answer is that for the most part, people operate under the desert-based notion of moral responsibility. This is the 'standard' notion of moral responsibility, and the one that people use the majority of the time. When asked about matters of praiseworthiness and blameworthiness, they are able to answer confidently most of the time. However, when determinism is made salient, when people are forced to confront the reality of human behavior and its causes, most are no longer able to respond as confidently as they were before. The tension is made obvious, and so the responses given are more split. After the problem fades in their minds, I suspect that the desert-based notion is once again the one that they make use of. It is, after all, the one that society seems to operate with.

So, to summarize, I believe that individuals, and society as a whole, operates using the desert-based notion of moral responsibility, an incompatibilist position. However, once the problem of determinism is made salient, intuitions shift in order to preserve the ability to assign praise and blame, and the individuals become more compatibilist. Once determinism is no longer at the front of their minds, they switch back to an incompatibilist viewpoint. The reason for switching back and forth is because the justification for praise and blame, which is important for any society, is based upon metaphysically untenable definitions of what is necessary for moral responsibility.

Honestly, I do not see this as being a major problem, despite the metaphysical failings, or rather metaphysical impossibility, of the desert-based notion. In fact, our tendency to think of moral responsibility in terms of desert probably does us more benefit than harm. When we think of ourselves and others as being responsible for our actions, we are less likely to attempt to excuse a refusal to change. That is how I see the desert-based notion of moral responsibility: as a catalyst for changing people's behavior. This is likely why that same notion has persisted so long

and so widely.

### Section 3

#### 3.1 Defense of the Current Conception/Why We May Not Need Change

Before I discuss alternate ways of regarding moral responsibility within the framework I have set up, I will take a few pages to consider in more depth a problem which I have already touched upon earlier. This is the problem that this conception of moral responsibility seems to make it okay to praise and blame someone for something they did not do, if it is beneficial to the moral system. This is related to another problem: this theory admits that metaphysically, no one can be responsible for their actions. I consider them closely related because I see no real difference between praising or blaming someone for something they cannot be metaphysically responsible for, and praising or blaming someone for something that they did not do. In neither case should a person be held responsible, metaphysically. So, how is it possible to justify doing so, regardless?

There are cases where moral responsibility has a use for moral systems, that much is certain. But simply because there is a reason, even a good reason, to do something does not mean that there are no other considerations that must be taken into account. What I have considered so far is whether systems of morality, as I conceive of them, are affected by the answer to the question of determinism. In particular, whether the role moral responsibility holds within systems of morality is affected by that answer. In order to properly explore this problem, it is necessary to consider not just the role moral responsibility plays in morality, but the general role morality plays in society.

It is likely that there are a wide variety of solutions that could be offered to this question,

but I have a particular strategy in mind that should be fruitful. I will make use of a similar system, already considered justified popularly, and attempt to draw parallels to systems of morality. To this end, for this section I will borrow a line of argument from Caruso. In “Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model” (2016), Caruso deals with a similar problem, specifically how to justify punishment of criminals given the thesis of determinism. The answer he arrives at is useful, I think, for clarifying my own position on how the function of moral responsibility in morality should be considered when deciding whether it is compatible with determinism instead of the desert-based notion. Particularly because Caruso deals with systems and how they are best implemented to achieve their goals, and why they are justified in often infringing upon the rights of individuals for the sake of others, as this paper treats morality as system similar to that of judicial law. It is important to note that I am not entirely unsympathetic to the concerns of individuals such as Galen Strawson when they claim that some kind of “self-determination” is necessary in order to be truly responsible for one's actions. It is intuitive, for reasons that I have already explained, that it is in some way unfair to blame or praise a person for an action that they could not help but take. It is because of this problem that I offer a different conception of what is necessary for moral responsibility.

The problem that Caruso is considering is very similar to the one that I am grappling with. However, instead of attempting to justify holding individuals morally responsible in spite of the thesis of determinism, Caruso assumes that the general problem with moral responsibility and determinism, the “could not have done otherwise” problem, is correct. Given this, what Caruso is trying to do is show that even if it is wrong to hold individuals responsible for their actions it is still possible to act correctly in administering certain punishments to criminals. After all, it seems that if determinism is true, punishing criminals for their actions is no longer

acceptable, leaving us without a means of dealing with criminal behavior. First, Caruso admits that it is impossible to apply retributive punishment if determinism is accepted. Retributive punishment derives its validity from the desert of the offender, and given that the individual could not have done otherwise than what they did do, it is impossible for them to deserve to be punished. But, according to Pereboom, who Caruso draws heavily from, there are other possible justifications for punishment that are less problematic for a free-will skeptic (2013). The discussed possibilities are for moral education, for deterrence, in self-defense, and for incapacitation. Pereboom believes that the most reasonable account is one that justifies incapacitation with the right to harm in self-defense. The first two means of justifying punishment are briefly considered, but ultimately discarding. Moral education because unlike children, adults generally understand the morality of the society they live in, and are less malleable in thought. Deterrence is rejected because it leads to morally unintuitive situations, such as justifying punishing the innocent, or administering punishments far harsher than seems fair, given the crime.

Instead, what Pereboom believes is the best option is to make an analogy between treatment of criminals and treatment of carriers of dangerous diseases. In the same way that a carrier of a disease is not in any way morally responsible for their condition, if the skeptics on free-will and moral responsibility are correct, neither are criminals. Regardless, we consider it acceptable to take certain actions which infringe upon the rights of infectious individuals, against their will if necessary, in order to act in both our interests and safety and their own. There is of course a caveat that their rights should not be infringed upon any further than is required in order to prevent the harm from being done. In the same way, it should be possible to justify infringing upon the rights of criminals using the same kinds of justifications. Given that the infringement is



no longer being done with the purpose to punish, the majority of policies we have in place currently would no longer be possible to justify under this new framework. Already, it should be possible to see the connection between what policy is being discussed here and the project I am undertaking. Once again, an assumption about morality needs to be made; that it is objectively true that certain acts are moral and others immoral, and that moral actions should be taken and immoral actions avoided. So, comparing the individual doing immoral things to someone who is infectious or criminal, it is justified to push moral responsibility upon someone that does not deserve it, both for the sake of their own moral development and for the well-being of others.

In fact, this is a justification, where minor infringements are made on individual rights for public benefit, we currently use already when assigning moral responsibility. Consider a young child, not yet developed enough to be justified in receiving praise and blame. We are aware of this lack of development, and of their relative blamelessness when they commit immoral actions. Yet regardless, we still praise them when they do things they should, and blame them when they do things they should not. We know they do not deserve to be blamed, but we still do it because the infringement of their rights, in this particular case their right to be blamed only when deserving of blame, is acceptable in order to help them develop morally. If we refused to offer praise or blame to children because they are not yet true moral agents, they would never become moral agents. This is, of course, not to say that this is the only possible way we could educate children to become moral agents.

Of course, another criticism is possible here. While it may be possible to justify moral blame when it is not deserved to individuals that are not moral agents in the interest of helping them become moral agents, how can the same justification be applied to adults? I think that for the majority of individuals, who on some level accepted the moral framework offered to them,

justification is possible. This is because it is not sufficient to simply praise and blame them until they accept the morality offered to them. In order for any system of morality to continue to exist within a society, it is necessary to actively maintain it. That is because, as said before, morality is not only concerned with the actions of individuals, but also their wills. In order for the individual to continue to trust in the system they are taking part in, they must continue to receive feedback on their actions from other members in order to confirm that that system is still extant.

Individuals are held morally responsible for their actions because they expect to be held responsible, and failure to meet this expectation could cause them to lose faith in that morality.

The major problem is in cases where the individual in question has completely opted out of the moral system being offered by their society. In such cases, I do not think it is justified to attempt to force responsibility for an action that they are not recognizing as being one of moral character, or as one with a different moral character than is being offered. Though for these individuals, the question of how they can function properly in a society with values so different from their own is reasonable to ask.

This section only addresses one of the potential problems someone could have with the current usage of moral responsibility in current moral systems. If there is some other aspect of this system that is taken as being a problem, the answers I give here may not be enough to give reason not to attempt a change. Despite this, I think that my response to this particular problem should be satisfactory, given that this is a line of justification that we make use of to a significant degree.

### 3.2 Why Change May be Difficult or Impossible

For the rest of my paper, I will concern myself with alternate conceptions of moral responsibility, and in particular, what challenges might arise from substituting in a new conception. P. F. Strawson makes an argument in his paper, *Freedom and Resentment* (1962) that the question of whether we should change our current systems is one that is ultimately pointless. Strawson begins the paper by detailing a particular argument that he makes reference to throughout: the argument that freedom can be defined in such a weak way that it does not conflict with the thesis of determinism. Strawson calls this the argument of the optimist. But what he calls the pessimist will not let this stand. There is a lacuna in this account, the pessimist says, in that this definition of freedom cannot support our use of praise and blame. And in addition, it is not as if this definition that the optimist offered is a controversial one. No one denies that freedom in such a weak sense is possible.

In order to explain why he thinks dwelling on these kinds of matters cannot be fruitful, Strawson first draws a line between two kinds of ways that we can regard others. In order to illustrate the difference between them, Strawson talks about our reactions when the actions of other individuals affect us. Specifically, when they do something that harms us. Now, when considering how to react to the other person's harmful actions, that is, what attitude we should take in regards to them, their intentions are very important, such as whether they are acting out of ignorance or good will, or out of indifference or malevolence. A minor harm can cause a great amount of resentment if it is done out of an indifference to our well-being, or if it was an intentional act on the other person's part. Likewise, considerable harm can be forgiven if the act was committed with ignorance in regards to the harmful effects, or if they were attempting to offer some kind of aid. Strawson considers this kind of consideration to be the “normal” way we

regard each other, and refers to them as participant reactive attitudes. The relationship we have with the other person is important as well. Friends, co-workers, family, lovers, strangers...depending on how close we are to them, we might adopt a completely different attitude for what is basically the same act and the same intentions that led to that act.

However, there are special sets of circumstances that cause us to evaluate other people differently. These are times when, as Strawson would put it, the individual is not “in their right mind”. These include times when the person is under great stress, or when they are forced to make a decision they would not usually make due to their circumstances. For these times, we suspend normal judgment, and take what Strawson calls the objective view. Other individuals that we take the objective view with are the mentally ill, who are also, in a way, forced by their circumstances to act in the way that they do. Children, because they are underdeveloped in various ways, including morally, also fit into this category.

While this objective viewpoint is something we are perfectly capable of doing, Strawson says that the participant attitude is the natural one for us to take, and while we are capable of adopting the objective attitude towards individuals that have none of the traits that usually cause us to adopt it, we cannot maintain it for very long. And here we come to the relevance of the thesis of determinism to what Strawson is doing. One might suppose that, if determinism is being considered, the question to be asked with whether to maintain the current system we have of engaging in the participant attitude, or whether we should adopt the objective attitude owing to the fact that an individual that is wholly determined can be said to be out of control of their actions in much the same way as the other individuals that Strawson says we use the objective attitude for. But Strawson considers this question to be a poor one. We do not really have a choice whether we use the participant attitude or objective attitude in our daily lives: we must

use the participant attitude. We are, as Strawson said before, psychologically incapable of maintaining the objective attitude for extended periods of time. So even if we do accept the thesis of determinism and the ramifications for the possibility of free will, we will still find ourselves unable to regard each other as this would seem to require. We will still resent others when they wrong us even if we know intellectually that, in a very real sense, it was impossible for them to think or do otherwise. Our society is based upon a variety of inter-personal relationships that require us to regard each other in the participant attitude. We are already thoroughly committed to the way we do things now, so much so that a wide-scale change to the objective attitude is effectively impossible at this point.

That is the answer Strawson gives to the question of how we should respond if the thesis of determinism is true: that the question is a pointless one to ask, given that we cannot change the way we do things regardless of what our intellect tells us is proper. The reason we are able to engage in the objective view with individuals such as the mentally ill is not because we accept some sort of theoretical grounding and act accordingly, but because we have accepted that having a normal inter-personal relationship with that person is impossible. That the behavior of such individuals is determined in any particular way is not a consideration.

At this point, Strawson turns to matters of morality. Taking the participant attitude we have, which regards the attitude we take to the actions of others that affect us, a moral view would be the same attitude, but generalized, or vicarious. That is, the proper attitude for individuals other than ourselves to take with regards to the actions of others, from the viewpoint of an uninvolved party. It is the fact that the attitude is vicarious that gives it its moral character. In addition, there is the proper attitude for oneself to take with regard to others. Thus, there are three types of attitude Strawson is concerned with: the demands of others with regards to oneself,

the demands of others with regards to others, and the demands of oneself in regards to others. We suspend these moral requirements for many of the same reasons that we suspend similar, non-moral attitudes. When the individual in question is unable, in some way, to understand, appreciate, engage in, or otherwise properly follow the moral requirements they have due to their circumstances. Once again, the question of whether, in some way, the individual's behavior is determined is not one that we ask when we are considering whether to suspend moral judgment.

After explaining his viewpoint on what separates moral and non-moral attitudes, Strawson once again asks the important question: if we accept the general thesis of determinism, does this mean that we should abandon our current system of moral attitudes to one that properly reflects that human behavior is determined? The answers he offers here are basically the same as the ones he offered for the question with regards to non-moral attitudes: that it is highly unlikely, regardless of whatever theoretical grounding we might accept intellectually, that we would be able to abandon a system that we are so deeply involved in. In the end, this is the answer Strawson gives to the entire debate on the matter of morality, free will, and determinism. The question is a pointless one, because whatever the answer is, we are too invested to accept any change.

Of course, this explanation is not without its problems. The first one is one which Strawson himself discusses, though he does not seem to think of it as a problem. Strawson says that in order for his explanation to work, it is necessary to abandon the metaphysics of morality.

“Because the optimist neglects or misconstrues these attitudes, the pessimist rightly claims to find a lacuna in his account. We can fill the lacuna for him. But in return we must demand of the pessimist a surrender of his metaphysics” (Strawson, 1962, pg.13).

But when Strawson mentions anything about 'metaphysics', he does not offer any explicit

definitions or explanations as to what he means. Therefore, I will attempt to provide a more robust account that hopefully captures Strawson's meaning when he uses the word within the framework of his paper. When he talks about the metaphysics of morality, I believe that Strawson is referring to realist moral facts, that certain actions are moral, and certain other actions are immoral, and these are as much facts as any fact about the physical motions of the universe. That is, they are true regardless of whether someone believes they are true, what opinion they have, and so on.

In order to accept Strawson's answer to the question of the relationship between determinism and morality, it is necessary that we abandon any belief that our moral approbation and condemnation are based upon realist moral facts just as these. Instead, what we are doing is simply engaging in a useful system, one which we are too invested in to seriously consider abandoning. And, considering that Strawson's answer to whether we should take the thesis of determinism into account in our moral system is that it is impossible for us to change, abandonment of the metaphysics of morality does seem necessary. After all, if there are such things as moral facts, then how we answer this question is extremely meaningful, and it is possible to be right or wrong, meaning that Strawson's response of saying we cannot change, so the answer does not matter does not seem adequate. So, anyone looking to solve the problem of determinism and morality while maintaining the metaphysics of morality will need to look elsewhere. In this way, it is similar to the viewpoint offered earlier, that morality should be viewed as a useful tool for society that we should continue to use, though in Strawson's case that we continue using it because we are incapable of not using it.

The other problem I see with Strawson's account is that it seems to encourage us to avoid trying to change the current system. While I believe that he is correct when he says that we are

heavily invested in the current system, so much so that any attempt to change would be incredibly difficult at the very least, does that mean that we should abandon any attempts to change? While I am not certain what a different society that took determinism into account would look like, it does not seem intuitively incorrect to say that we would not need to abandon normal inter-personal relationships in order to have a society that more accurately reflects the level of free will that humanity actually has. And though it may be a slow and difficult process to attain a different kind of society, if such a change is worthwhile, we should strive toward it, regardless. And it is not altogether clear that Strawson is entirely correct in his assertion that it would be impossible, or difficult, to change the way we conduct inter-personal relationships. The fact that our relationships are the way they are now is entirely contingent, with there being no logical inconsistency or known physical law preventing a different kind. Strawson believes that humans, psychologically, are incapable of changing our attitudes to an objective one, even if determinism becomes a well-accepted truth, but he does not say this with absolute certainty. Whether or not we actually can and whether we should are different matters. But if we can, and if it would benefit us to do so, it would not be unreasonable to try. Answers to questions of morality are very important to us as agents, free or not free, and Strawson's assumptions do us no favors in generating those answers. Still, Strawson is important for what I am doing because his explanations on why we are so attached to desert-based notions of responsibility are my basis for claiming that that view, however flawed Galen Strawson and others show it to be, has an important place in our society, and how our current morality functions. So not only would Strawson say that any attempt to change how we define moral responsibility is doomed to fail, but that any such attempt regarding our embedded systems is. Therefore, regarding the alternate conceptions of moral responsibility offered below, Strawson would likely consider them a waste



of time. While I do not fully agree with Strawson on this point, the arguments he makes are reasonable, and something that should be taken into consideration when endorsing an alternate conception.

### 3.3 Potential Problems with Alternate Conceptions

While I think there can be a variety of potential systems that could work in place of the one which I have described, if Strawson is incorrect in his assertion that such a change is not possible, I believe that there are at least two major pitfalls one must avoid when theorizing about these matters. That is, when making a theory of moral responsibility, it is necessary to avoid both of these problems in order to match or exceed the current system in terms of usefulness.

In order to explain the first of these pitfalls, I will turn to J. J. C. Smart and his work “Free Will, Praise, and Blame” (1961). Smart also represents a kind of compatibilist in that he believes that it is possible to be responsible, or at least hold others responsible, even if determinism is true. Given that I believe that moral responsibility, though also of a different kind than usual, is compatible with determinism, I too could be said to be a compatibilist. Smart spends a considerable amount of time in this paper attempting to show that the “free will” that libertarians seek is meaningless, and for my part I agree with him. I will not go into his arguments as the arguments that Galen Strawson makes later against the idea of a desert-based moral responsibility are significantly easier to understand and represent the most basic argument against such a conception.

The part which I am interested in comes afterword, where Smart discusses praise, dispraise, and blame, and what function they fulfill in our society. Smart considers the case of two schoolboys, both which failed to complete the homework assigned to them. The first boy

failed to do so because he is, unfortunately, quite stupid. Smart says that in this case, that it is wise for the teacher to refrain from punishing the boy. No matter what threat of punishment or actual punishment is delivered, the boy is simply incapable of completing the task given to him. He is not to blame because he could not have done otherwise, by which Smart means that even if the circumstances were different, the outcome would remain the same.

The second boy also fails to complete his homework, but not because he is unable to, like the first boy, but because he is lazy. In this case the teacher should punish the boy, and tell him that he could have done it. Not to say that the boy's behavior is not the product of his genes, upbringing, and environment, because Smart would certainly say that it is. But part of the boy's environment is the teacher, and whether or not he does his homework can be influenced by the reaction the teacher gives to him not doing his homework.

So, according to Smart, there is a pragmatic justification in making use of threats, punishments, rewards, and ascription of responsibility even if determinism is assumed to be true, given that these things are capable of influencing people's behavior. He goes on to claim that the very fact that we make such heavy use of these things implies a certain level of consent to the idea that human behavior is largely determined.

Smart then turns to the concepts of praise and blame. Praise can be used in two ways. One, it can be used as an opposite to blame. A person is praised for doing well, and blamed for failing to do well. But we also praise people for things that they, in no sense, have any control over such as what talents they were born with. So, while we will praise a person for having a talent for art, will do not blame anyone if they are born without any such talent. We may call a person ugly, or a soccer player uncoordinated, but we are not blaming them for these traits. When praise and dispraise is used in this way, we are simply grading the trait in the person as either

good or bad. Smart believes that this grading has the same function in our society as our methods of grading other things. We use it in order to inform on the quality of the individual in question, to indicate what they are like. Smart considers this to be the primary function of praise and dispraise, with influencing individuals to take or avoid certain classes of action being a secondary but still important function.

After this, Smart goes back to the first function of praise, as an opposite to blame. The difference between praise and dispraise and praise and blame is that when we use praise and blame, we are not only describing the quality of the individual and their traits, but we are ascribing responsibility for that quality. Blame used in this sense can be a dispassionate observation, but often is not. Smart believes that the average person's viewpoint on ascribing responsibility is that, if a person's actions were based upon their genetics and environment, it would be unfair to praise and blame. This is because they hold to a metaphysical viewpoint on free will that Smart considers incoherent, as he argued in the first part of his paper. Because they believe in this incoherent concept of free will, people do not just grade other people through praise and dispraise, but judge them through praise and blame, and this is what Smart has a problem with. Smart believes that it is necessary to abandon this metaphysical viewpoint and act in accordance with the utility of actions. As far as Smart can tell, judging is an action that does not have a meaningful effect upon a person's behavior, at least compared to the myriad other functions which influence it, and is therefore lacking in utility. Therefore, as praise and dispraise has utility, and praise and blame lacks it, we should use praise and dispraise.

Smart believes that when we judge other people instead of merely grading them, we are taking an irrational action. And while it may be true that in the way Smart described it, we are making a mistake, this is not only a natural mistake to make, but one that is built into the current

systems of morality that we current have, as I have already shown.

Smart believed that the main reason judging instead of grading is irrational is that the effect of the judgment on another person's behavior is fairly minor compared to all the other factors that are affecting them. I think that this is a mistake. The pressure to conform to the judgment of one's peers is well-known to be incredibly powerful, particularly if that judgment comes from an individual that is recognized as being competent to judge. While the judgments of individuals that we are not familiar with, or who we have graded as not being fit to make such judgments, may have little effect on a person's behavior, as Smart claims, anyone that has been judged competent, or even someone whose judgment is of unknown quality, are more likely than not to cause a change in the behavior of the person they are judging. Though of course, this depends on how susceptible the person in question is to being influenced in such a way, humans are social creatures, and the drive to conform to the desires and expectations of other members of one's community is a definite influence on human behavior.

Smart also criticized the way people make their judgments, claiming that they fail to be properly dispassionate in making them. While this seems to be a problem, and normally I would agree that it is, in the case of assigning moral responsibility, the passionate and occasionally irrational nature of the judgments is an indicator that the individual agrees with and believes the content of the moral system. Recall that morality is not only concerned with constraining the actions of individual's, but also with constraining their will. An action is in agreement with morality if and only if the action undertaken is the correct one, and if the action is undertaken for the correct reasons, with the correct will. To have a moral will requires that the individual in question ascribes to and agrees with, if not fully than significantly, the code that the morality posits. And when a moral code is believed in, as is required in order to have a moral will, it is

only natural that judgments made by such individuals would fail to be dispassionate. So, while it would be ideal if people could judge others, or merely grade them, in the way that Smart wishes, this is simply not the way that things are now. If it were possible to change our behavior in the way that Smart believes we should, that would perhaps be for the best, but I am uninterested in advocating or condemning calls to change the current systems of morality. Though later on, I will look at other possible interpretations of the current systems to see if those systems can fit particular requirements we might have of them.

However, we may have reason to consider Smart's viewpoint as being insufficient if we were to adopt it. For Smart, the purpose of praise and dispraise is to inform others about how they are graded. When it acts as a motivator for change, it is because whatever is being encouraged or discouraged is already recognized as something worthwhile. In the case of the lazy boy that does not do his homework, he already knows that it is in his best interests to do his homework. So, in those cases, and similar ones, Smart is correct to say that praise and dispraise are sufficient. And in many of the tenets that systems of morality commonly posit, they are quite similar, so the assumption that when we want people to avoid doing some things and actively pursue some others, praising them when they act properly and dispraising them when they act improperly seems like it would be enough, at least on the surface. But Smart only talks in terms of the utility of actions, and what is considered moral and what provides the most utility, while occasionally asking the same things of us, are fundamentally different. Or at the very least, there is nothing preventing these two metrics from offering different judgments on what action to take.

There are times when morality asks of us things that would seem difficult to justify as a rational action. Take the prohibition on lying, for example. There are times when telling the truth is the most rational action to take, but we are not told to tell the truth only when it is what brings

about the most good, but to always tell the truth, often without regards to the consequences.

Because morality is so often absolutist, merely grading them or explaining why the moral rule is in place may not be enough. Smart assumes that we can rationally convince others to do what is best for them, but what is moral and what is best, or even good or beneficial, for someone need not be the same thing. This is particularly salient in cases where self-sacrifice is required of us; how could praise and dispraise convince anyone to act in such a way?

In order to convince people to behave in such ways as morality often requires, a different approach is commonly used. We appeal to things other than to rationality. We do not use evidence to convince others that something should or should not be done, but rather we use their desire for praise and aversion to blame. When we offer praise and blame, we instill in the person a moral belief. For example, if society tells a person repeatedly that stealing is wrong, praises them for not doing so when they have the opportunity, and blaming them when they do steal, that person will likely avoid stealing. They will also come to believe stealing is wrong in order to justify both their own behavior and to justify and understand why society is telling that stealing is wrong. This moral belief, rather than the desire for praise and aversion to blame, becomes the motivation for undertaking moral action. This belief is not rational because, while in some sense there is evidence in that this is what society is telling you is true, it is the desire for praise, aversion to blame, and other similar emotions that first motivate. That is not to say that moral belief cannot arise out of rational concerns. Someone may be against stealing in general because they desire not to be stolen from more than they wish to steal. But in cases where praise and blame act as a primary motivator, these kinds of concerns do not need to be considered.

Of course, it is possible to make an argument that these kinds of tenets are not the kinds that should be encouraged, but what morality should be is not of concern here, merely what it is.

And it is an absolute fact that these kinds of tenets are common. Now, it may be the case that even if my arguments against Smart are accepted, there are other ways of achieving the exact same result. This is not a problem, as I do not wish to do the same thing that I criticize Strawson for doing, promoting inaction. If a better way of getting people to act morally can be found and implemented, there would be no contradiction with my account. I merely wish to explain that Smart's assertion that we should not engage in praise and blame is uncalled for, given that praise and blame are an important part of how we assign moral responsibility in current systems.

Still, Smart may very well say that the methodology he offers is superior to the one which I described just now. My defense of the current system was not intended to be an argument for why it should be accepted over all other systems, but an explanation for how the current system functions and why it may not be necessary to replace it with a different one. Therefore, Smart could claim that while the system we have now is functional, a different one would perform the exact same job, but do it better. While, in a way, Smart might be correct that his system would be better than the current one, there is an important point to keep in mind that may offer reason to prefer how we do things now. It is at this point that I wish to explain the first of the two major ways a system of moral responsibility can fail, using Smart's methods as an example.

Let us recall the primary change that Smart wishes us to make to our current system of moral responsibility: that instead of using emotional praise and blame, we should instead use dispassionate praise and dispraise. His argument is that the second system would simply perform better than the first. But here is the problem: I think that Smart has failed to consider an important facet of any moral system. That facet being that moral systems need to be applied, for the most part, very widely. Because moral systems need to be applied widely, the range of individuals that need to be able to successfully engage with the system and obtain useful results

is very wide as well. This range of individuals encompasses a range of capacities, with some being capable of making use of highly advanced and complicated systems, and others struggling to make use of very simple ones. To speak bluntly, some people are smarter than others, and a moral system needs to be usable even to the least capable members.

This is the problem that I see with Smart's proposal: I am doubtful that those least capable members would be able to make good use of a system such as the one Smart thinks should be implemented. Therefore, the pitfall that Smart has committed is the one of over-intellectualizing the system of moral responsibility to the point where the ability of that system to be applied as widely as it needs to in order to function is cast into doubt. That is not to say that among those that are capable of using this system, it would not function better than the current one, as Smart claims. So perhaps it would be reasonable for those individuals to change their use of praise and blame to praise and dispraise. For the majority, however, better results would likely be gained from keeping the current system.

### 3.4 Another Problem

The second potential problem with an alternate conception of what is necessary for moral responsibility is best illustrated by Frankfurt's "Freedom of the Will and the Concept of a Person" (1971). While I do agree with Frankfurt in his rejection of the conception of moral responsibility that Galen Strawson offers, that metaphysical responsibility is necessary for moral responsibility, I think that the conception he offers in place of it has a very important problem.

The notion of will is often used to describe the first-order desires of an individual, whatever reasons motivate them to act. They take the form of sentences such as "A wants to X" with X being an action of some kind. Frankfurt's definition of will is the effective desires, those



that move a person all the way to action. There are also second-order desires, which take the form of “A wants to X” with X being some first-order desire. So basically, “A wants to want to X”. Second-order desires are important because Frankfurt believes that they are necessary in order to be a person, as he defines it. But not only second-order desires, but what he calls second-order volitions. Second-order volitions are desires where the individual desires for a certain desire to be their will. That is, they desire for a certain desire to move them to action.

Anyone without second-order volitions cannot be a person, and Frankfurt calls all such entities “wantons” (Frankfurt 1971 pg.11). Wantons do not care about their will, and do not, or cannot, be concerned with the content of their desires. While it is possible for them to apply considerable reasoning ability in determining how to best get what they want, a wanton does not care which of their desires actually moves them to action, only which of the conflicting desires is strongest. How desirable their desires are is something they do not concern themselves with. Wantons include things such as all non-human animals, children, and possibly even some adults.

Frankfurt illustrates the difference between a person and a wanton with a story of two drug addicts. Both addicts have the same physiological condition regarding their addiction, and both succumb to their addiction periodically. One of them hates their addiction, and struggles, though in vain, to rid themselves of it. This unwilling addict has conflicting first-order desires in that he both desires to take the drug and desires to not take the drug. In addition, he has a second-order volition: he wishes for his desire not to take the drug to be the desire that moves him to action, the desire that constitutes his will. The other addict does not have any second-order volitions. He does not care which of his first-order desires moves him to action. While he may experience difficulties in obtaining the drug or conflicts between his first-order desires, he has no preference for how these conflicts are resolved. It is not even accurate to say that he is neutral

with regards to first-order conflicts, given that his entire identity simply is his first-order desires.

It makes a difference to the first addict which of his first-order desires is acted upon. He wishes for his desire to take the drug to go unfulfilled, and he wishes for his desire to not take the drug to move him to action. It is this identifying with one desire and rejecting another that makes it meaningful to talk about free will. It is important to note that the language Frankfurt uses here is easily confused, for he talks about a will that is free and free will as two separate concepts/capacities. A will that is free is a will that the individual is capable of determining for themselves. That is, that the individual is capable of determining what desire it is that moves them to action. An individual that has free will is different. What it means to have free will is that the individual's will that they wish to have. So, if the desires that motivate them to action are the one/ones they wish to motivate them, then they are acting with free will. It is therefore possible that someone not have a will that is free while still acting out of free will.

Frankfurt illustrates this with another example of a drug addict. This third addict is the same physiologically as the other two, but unlike them he has a second-order volition to be moved by his desire to take the drug. His will is not free because the desire that will move him to action will be the desire to take the drug regardless of what his second-order volitions are. But he is still acting with free will because the desire he wishes to move him to action is the desire that is actually moving him to action. This is the definition of free will that Frankfurt settles on, and what he believes is necessary in order to have moral responsibility. The requirement that we have free will in the strong sense seems to push in the direction of human actions being in some way indeterminate, but this does not seem to get responsibility, either. In fact, if our actions are indeterminate, I would be inclined to say that we would be less responsible for our actions in any sense than is determinism were true. A person that is acting randomly, as this seems to imply, is

not even making choices, let alone free ones. So why this would be held as a necessity for free will is a mystery to me. Libertarians would likely claim that being indeterminate does not mean being random, or that not being determinate does not mean being indeterminate, but I have yet to find a reasonable explanation for the supposed middle ground between determinism and indeterminism that incompatibilists seem to be advocating. Moral responsibility, under this conception of what it means to have moral responsibility, will be limited to individuals that are capable of having second-order volitions, and they will be responsible for a particular action if and only if they had a second-order volition to have the desire that motivated that action constitute their will. Naturally, in cases of obvious external constraint, where the individual in question is physically unable to act other than how they did, rather than simply being unable to choose otherwise, will also not be held responsible for their actions.

But while this sounds like a good way to define what is necessary for moral responsibility, like Smart, there is a problem. This problem comes from a conflict between the role of moral responsibility and the criteria for moral responsibility that Frankfurt offers. If I am correct, what moral responsibility does is acts as a motivator to cause people to follow the rules of a moral system. Through actions such as praise and blame, acting in accordance with the morality the public accepts is enforced. The problem being, I do not think that moral responsibility is capable of fulfilling this role if we accept what Frankfurt tells us. Frankfurt says that the difference between someone who is and someone who is not moral responsibility for their actions is a difference in second-order desires: what they want to want. But if this is the case, how are other individuals, who do not have access to the content of anyone else's desires, second-order or otherwise, to determine whether to hold someone as moral responsible or not? If moral responsibility is to play this role, other people need to be able to tell when someone else is

or is not moral responsible, and given Frankfurt views on moral responsibility, this would be impossible. In fact, I think it may be doubtful whether even the person in question can tell with certainty whether they should be held moral responsible. Self-reporting on the content of one's own mind is a notoriously unreliable process, so even if someone is being completely honest, they may not be able to tell whether or not they have some second-order desire. So once again, the problem is one of practicality: I do not think Frankfurt's definition of moral responsibility would work in practice, given the challenges above.

This pitfall, along with the one illustrated by Smart's views, are ones that need to be avoided if a properly functioning moral system, where moral responsibility is able to perform its necessary goals, is to be had. So, it may be the case that, rather than try to substitute our current working definition of moral responsibility for a new one, we should instead be happy that the current system works as well as it does. And if we do manage to discover a new way of doing things that both works better than the current system and that we can capable of implementing, so much the better.

## Conclusion

If we look at the way morality works, it is not significantly different from a wide variety of other systems that people are familiar with. It is concerned with having certain things done and certain other things avoided, all while requiring a certain frame of mind. Moral responsibility, through actions such as praise and blame, is an incredibly powerful tool for changing behavior that morality has available to it, regardless of the way that it is used. I have hopefully shown that when holding individuals morally responsible for their actions, even if the thesis of determinism is true, moral responsibility is still perfectly able to fulfill the role it has in

moral systems. Despite the metaphysical problems that the desert-based view has, our current morality seems to function quite well, in terms of whether it can cause the individuals under it to adhere to its tenets. Of course, the tension between the metaphysical problems of the desert-based conception and its usefulness in moral systems is a possible cause of the entire debate between moral responsibility and determinism, though I suspect there are other factors in play as well.

If we do wish to change the way we utilize morality, it will likely take significant effort given how embedded the current way of doing things is. We would also need to make sure to avoid the pitfalls I have described, and likely others besides.

I have described how I see current moral systems as functioning in order to show that this particular conception has no problem with determinism, but there is no reason to believe that it is impossible to change the way we do things for something a little better, despite what Strawson may say on the matter.

#### Works Cited

Gregg D. Caruso, "Free Will Skepticism and Criminal Behavior: A Public Health-Quarantine Model" *Southwest Philosophy Review* 32 (1): 25-48. 2016

Chisholm, "Freedom and Action," in Lehrer, ed. (1966)

Harry G. Frankfurt, "Alternate Possibilities and Moral Responsibility" *The Journal of Philosophy*, Vol. 66, No. 23. (Dec. 4, 1969), pp. 829-839.

Harry G. Frankfurt, "Freedom of the Will and the Concept of a Person" *The Journal of Philosophy*, Vol. 68, No. 1 (Jan. 14, 1971), pp. 5-20

Robert Kane, "Two Kinds of Incompatibilism" *Philosophy and Phenomenological Research*, Vol. 50, No. 2 (Dec., 1989), pp. 219-254

Eddy Nahmias, Stephen Morris, Thomas Nadelhoffer, and Jason Turner, "Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility" *Philosophical Psychology* Vol. 18, No. 5, October 2005, pp. 561–584

Derk Pereboom, "Free Will Skepticism and Criminal Punishment" In Thomas Nadelhoffer (ed.), *The Future of Punishment* (pp. 49-78). New York: Oxford University Press.

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

Galen Strawson, "The Impossibility of Moral Responsibility" *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*, Vol. 75, No. 1/2, Free Will, Determinism, and Moral Responsibility (Aug., 1994), pp.5-24

J. J. C. Smart, "Free Will, Praise, and Blame" *Mind*, New Series, Vol. 70, No. 279 (Jul., 1961), pp. 291-306