

Grounding Responsibility:  
Helen Steward's Libertarianism and a Hemi-Incompatibilist Alternative

by

Mary Renaud

A thesis submitted to the Faculty of Graduate and Postdoctoral Affairs in partial  
fulfillment of the requirements for the degree of

Master of Arts

in

Philosophy

Carleton University  
Ottawa, Ontario

© 2013, Mary Renaud

## **Abstract**

This project has two purposes. The first is to evaluate Helen Steward's libertarian account, as presented in *A Metaphysics for Freedom*, and assess its ability to help ground moral responsibility. The second is to provide an alternative, Hemi-Incompatibilist account of a variety of responsibility that is compatible with the denial of agent causal control.

An overview of a selection of accounts from the moral responsibility debate is provided, followed by a discussion of Helen Steward's account specifically. I then present my Hemi-Incompatibilist account, followed by the brief investigation of two similar views. I conclude that Helen Steward's libertarian account does not succeed in lending support to those hoping to ground a deserts-based variety of responsibility. I show that my Hemi-Incompatibilist account grounds another variety of responsibility, responsibility qua onus, that supports normative prescriptions and proscriptions despite the denial that agents possess regulative control.

## **Acknowledgements**

I have many people to thank for the completion of this project and the studies leading up to it. Some of the content in this thesis has emerged from papers developed in seminars and tutorials with Professors Gordon Davis, Andrew Brook, and David Matheson. The feedback received from these submissions was invaluable to the development of my thesis project.

I would like to express my sincere appreciation to Professor Andrew Brook for leading me through the reading of Helen Steward's account and for helping to illuminate the challenging material. Without his guidance, patience, and warm encouragement, I would not have had the opportunity to study this most interesting account with which I wrestle in this project.

I owe an important debt to Professor David Matheson for pushing me on assumptions I was taking for granted during the Research Seminar. These discussions led to the development of a more mature version of the view I begin to outline in this project. I would also like to thank him for his advice on honing philosophical skills, which has proved invaluable, and his patience with my many questions on the subject. I would like to thank Gabriele Contessa, as well, for key elucidatory discussions about metaphysics that opened my mind to additional considerations about the material just when they were most needed.

I am also deeply grateful to Carleton University's Department of Philosophy's faculty as a whole for five years of support and instruction from professors who have been knowledgeable, interested, engaged, and engaging. Special thanks, as well, to Sandra Kirkpatrick, our departmental administrator, for

her constant support, calm, sense of humour, and tolerance in the face of my many administrative blunders.

I have greatly benefited from interesting exchanges with my fellow postgraduate students on philosophical topics of all flavours. I would like to thank Jonathan Courtney, in particular, for his friendship, pleasant company, and the many heated and humourous debates that compelled me to better develop my arguments.

I want to thank Derk Pereboom for permission to work with his draft. I would also like to express my gratitude to Carleton University, the Social Science and Humanities Research Council, and the Ontario Graduate Scholarship program for the generous financial support that allowed me to devote my attention to my studies.

Finally, I would like to express my deepest appreciation to Professor Gordon Davis for his selfless generosity with his time, for his astute, constructive, and insightful comments and feedback, and for his inexhaustible support. I have benefited immeasurably from our extended discussions and from his entertaining company, interesting seminars, and sound advice.

## Table of Contents

<b>ABSTRACT</b> .....	<b>II</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>III</b>
<b>TABLE OF CONTENTS</b> .....	<b>V</b>
<b>1. INTRODUCTION</b> .....	<b>1</b>
<b>2. THE DEBATE ON MORAL RESPONSIBILITY</b> .....	<b>4</b>
SEPARATING FREE WILL AND MORAL RESPONSIBILITY .....	4
DETERMINISM, INDETERMINISM, AND FATALISM .....	5
CLASSIFYING THE VARIOUS ACCOUNTS OF MORAL RESPONSIBILITY .....	7
AGENT-INTERNAL, AGENT-EXTERNAL, AND INCLUDED-AGENT ACCOUNTS .....	8
AUTONOMOUS AND ANCHORED ACCOUNTS .....	8
INTRODUCTION TO THE SPECIES OF VIEWS .....	10
INCOMPATIBILISM.....	11
LIBERTARIANISM.....	12
HARD DETERMINISM .....	13
HARD INCOMPATIBILISM .....	14
DERK PEREBOOM’S HARD INCOMPATIBILISM.....	15
GALEN STRAWSON’S BASIC ARGUMENT .....	16
CLASSIFYING INCOMPATIBILISM.....	17
COMPATIBILISM .....	17
CLASSICAL COMPATIBILISM .....	18
SECOND-ORDER DESIRES AND REACTIVE ATTITUDES .....	19
CLASSIFYING COMPATIBILIST ACCOUNTS.....	21
SEMI-COMPATIBILISM .....	22
FISCHER’S SEMICOMPATIBILISM .....	22
<b>3. HELEN STEWARD’S LIBERTARIANISM</b> .....	<b>26</b>
AGENCY INCOMPATIBILISM .....	26
SETTLING .....	26
AGENTS.....	27
ACTION.....	30
REFRAINMENT .....	31
JOE’S DECISION .....	32
ONTOLOGICAL PLURALISM ABOUT CAUSES .....	34
TOP-DOWN CAUSATION.....	35
PROBLEMS WITH STEWARD’S ACCOUNT: A UNIQUE KIND OF SETTLING? .....	36
MORE PROBLEMS IN STEWARD’S ACCOUNT: “SETTLED WHEN” VS. “SETTLED BY” ..	41
PHYSICAL REALIZATIONS.....	43
TWO-WAY POWER OF REFRAINMENT: AN INADEQUATE DEFENCE .....	47
GROUNDING RESPONSIBILITY .....	53

<b>4. RESPONSIBILITY IN THE SENSE OF ONUS.....</b>	<b>55</b>
RESPONSIBILITY AS A DIVISIBLE CONCEPT .....	55
WHAT OUGHT I TO DO? .....	56
A VARIETY OF SENSES OF RESPONSIBILITY .....	58
“‘OUGHT’ IMPLIES ‘CAN’”, THE VARIETY OF OUGHTS, AND THE UNKNOWABLE FUTURE .....	60
“OUGHT TO” IMPLIES “CAN, FOR ALL I KNOW” .....	64
THE ROLE OF EVIDENCE .....	65
APPROPRIATE ONUS-IMPLYING OUGHT CLAIMS.....	67
ONUS REMAINS UNHARMED BY THE LACK OF FREE WILL .....	68
RESPONSIBILITY HEMI-INCOMPATIBILISM .....	69
<b>5. SIMILAR VIEWS .....</b>	<b>71</b>
BRUCE WALLER, <i>AGAINST MORAL RESPONSIBILITY</i> .....	71
WALLER’S SENSES OF “OUGHT” .....	71
WALLER AND THE REJECTION OF “‘OUGHT’ IMPLIES ‘CAN’” .....	74
USEFULNESS .....	77
AGAINST <i>SOME KINDS</i> OF RESPONSIBILITY .....	78
DERK PEREBOOM, “FREE WILL SKEPTICISM, BLAME, AND OBLIGATION” .....	78
OUGHTS OF AXIOLOGICAL EVALUATION AND OUGHTS OF SPECIFIC AGENT DEMAND .....	79
OUGHTS OF AXIOLOGICAL RECOMMENDATION .....	80
<b>6. CONCLUSION .....</b>	<b>82</b>
<b>WORKS CITED .....</b>	<b>83</b>

## 1. Introduction

The inquiry into the grounds for ascribing moral responsibility to agents is an important one, as it has upshots that reach a number of other fields. Questions of responsibility affect our legal and criminal practices, our moral education, as well as our day-to-day personal and relational experiences.<sup>1</sup> Many assume that most, if not all of their fellow humans have some kind of free will and are morally responsible for their behaviour, at least in most circumstances. This assumption, however, deserves to be examined closely. If we are mistaken, it is likely that we are engaging in a number of behaviours, especially retributive behaviours, that we would consider unacceptable, perhaps even unethical.<sup>2</sup>

In this work, I will examine Helen Steward's account of Agency Incompatibilism as an approach for defending the kind of incompatibilist freedom for which some might hope and that might vindicate a belief in the type of agency that appears to lie at the basis of standard notions of moral responsibility. I will also present an account of my own that, although it does not endorse the sort of freedom for which Steward and many other libertarians might hope, shows that certain normative practices and varieties of responsibility are well insulated from the denial of free will.

As we will see, Steward's account of agency would not be enough to silence determinists. The latter can accept her view as coherent and yet still deny that there exists anything that resembles the kind of agent or the variety of action

---

<sup>1</sup> The term "responsibility" will be intended to refer to moral responsibility throughout, unless otherwise specified.

<sup>2</sup> This depends, of course, on the normative theory to which one subscribes.

that Steward has described. Steward's account could, however, make plausible a naturalistic explanation of incompatibilist agent causation and, if she is correct in supposing that this could provide the basis for a metaphysics of freedom, one without ontologically suspicious powers or states<sup>3</sup>, it could offer the incompatibilist a new, more robust, physicalist libertarian option<sup>4</sup>. Furthermore, if Steward can defend the kind of agent causation that makes it such that the agent's activities are, in a strong sense, *up to* the agent, it might pose very serious problems for the hard incompatibilist as the latter claims that freedom and the variety of agency it requires are incompatible with both determinism and indeterminism. Nevertheless, the stalemate with the determinists would remain.

I will begin by clarifying the terms of the discussion. The concepts connected to this inquiry have been conceived in a number of different ways. I will discuss some of these ways and will make clear the senses of each that I will be using throughout this work. This will lead into a brief history of the lengthy debate and the various positions, including the hard incompatibilist view to which I subscribe. As I will later show, the latter is the view most challenged by Steward's account. I will then discuss why I believe compatibilism does not provide what many seek in the discussion about freedom and responsibility.

Following this I will provide an exposition of the account presented by Helen Steward on the nature of agency and its incompatibility with determinism in

---

<sup>3</sup> Throughout, I am using "metaphysics" in the sense that refers to the philosophical discipline that deals with topics such as ontology and causation, and not in the sense meant to imply a disconnection from reality. Metaphysically suspicious powers are suspicious not because they are metaphysical, but because their ontological foundation is suspect.

<sup>4</sup> The libertarian views discussed herein are not related to political libertarianism. Libertarianism in the context of free will claims that free will exists and that universal determinism is false. This will be explained below.

her recent book, *A Metaphysics for Freedom*. This account presents what I believe to be the most plausible form of libertarianism and the most rigorous challenge to the hard incompatibilist position, which I find the most compelling position overall. I will then attempt to evaluate the view and discuss its viability as well as its implications for moral responsibility. Ultimately, I will argue that it does not succeed in presenting a clear and tenable account of indeterministic agency and that, if it had succeeded in providing such an account, it would not have been one capable of grounding moral responsibility.<sup>5</sup>

Next, I will turn my attention to the presentation of a revisionist, hemi-incompatibilist account of moral responsibility that is compatible with the tenets of hard incompatibilism about free will. This account relinquishes merit-based praise and blame,<sup>6</sup> but attempts to show that the responsibility to follow certain proscriptions and prescriptions remains, even without the kinds of freedom advocated by the traditional incompatibilist. Following this, I will discuss, briefly, two views that are similar to, but contradict my account, and will attempt to show that these views do not present a challenge for my own.

It is my hope that this work will show the need for further clarity in the debate, the reasons the current views are not sufficient, and the tenability of my hemi-incompatibilist account of moral responsibility.

---

<sup>5</sup> Responsibility of a non-compatibilist variety.

<sup>6</sup> Those of any strong variety, in any case.

## 2. The Debate on Moral Responsibility

### Separating Free Will and Moral Responsibility

As is the case with many philosophical concepts, the task of providing a definition of "free will" is not an easy one. Some argue that free will requires contra-causal control, sometimes called "regulative control". This is a kind of control over one's own actions that allows the agent of those actions to be a causer, whose causings are not themselves caused (at least not necessarily) and who has, therefore, genuine alternative possibilities open to her (Fischer, My Way 6). Others maintain that it requires only that we be able to act without interference from physical, psychological, social, or other constraints that hinder upon our ability to act as we choose. Still others believe that this freedom requires something like the ability to act in accordance with one's authentic character or second-order desires. Whatever the definition, it is clear that free will and moral responsibility are closely linked; though it remains unclear in which direction the connection flows. There are those who argue that one is responsible only if one has free will. There are others who believe the explanation flows in the reverse direction and that one has free will only if one is responsible.

I will bypass the question of the definition of freedom altogether and will focus only on moral responsibility. Given that accounts of the criteria for moral responsibility and its appropriate attribution are at least as contentious as are the definitions of free will, it would be too large a task to tackle both in a single work, and so I must restrict myself solely to questions about responsibility.

For reasons that will become evident below, it should be made clear that during the first portion of this work I am referring, when I refer to moral responsibility, to a kind of responsibility that is connected to praiseworthiness and blameworthiness. Whether the connection is because we are blameworthy only if we are responsible, or we are responsible only if we are blameworthy is unimportant. What is important is that the type of responsibility to which I am currently referring is one in which one of these is required for the other. I will introduce another sort of moral responsibility below but, for the sake of clarity, it is best that I not introduce it prematurely.

### Determinism, Indeterminism, and Fatalism

Another concept important to the debate is that of determinism, and what will interest us here is universal determinism, specifically. Universal determinism affirms that the state of the world and physical laws at a given time-slice necessitate the state of the world at all subsequent time slices.<sup>7</sup> It holds that future events are necessitated by prior events, in combination with the physical laws. The state of the world at time  $t_2$  is produced, necessarily, by the laws of physics and state of the world at time  $t_1$ . The state of the world at time  $t_3$  then follows necessarily from the state of affairs at time  $t_2$ , and so on. As a result, an entity with a god's-eye-view, perfect knowledge of the laws and current conditions (and presumably unlimited computational power) should be able to predict all future states of the world. Regardless of our or anything else's ability to predict them, all future states are already determined to come about and their

---

<sup>7</sup> Throughout the remainder of this project, all references to "determinism" refer to universal determinism.

details already fixed, according to this view. Following the initiation of an initial causally deterministic series (assuming there was such a happening), all future events and states are determined, even in advance of their occurrence. There is but one possible way the future can unfold.

It is important to differentiate determinism from fatalism. Fatalism is the view that no matter what we do, there are certain events that are unavoidable. It is the belief that no matter what actions one takes, certain results cannot be avoided (Kane 19-20). Someone who believes in fatalism may believe, for example, that they were "meant" to meet someone in particular on a specific day at a specific time and that, no matter what actions they took before that time, the meeting was destined to occur. A fatalist can believe there are many possible routes to and between these predestined events, whereas a determinist believes there is only one possible way in which the future can unfold.

There are ways to think of fatalism that make determinists appear merely to be a variety of fatalist; one that believes that *all* events, rather than just some events, are destined to be as they are. The terms are best kept separated, however, as fatalism often implies that there is room for indeterminacy in between predestined events. Additionally, because the compatibilist and the incompatibilist will view the phrase "no matter what we do" so differently, if we were to think of determinism as a brand of fatalism, we would need to refer separately to "compatibilist fatalism" and "incompatibilist fatalism" in order to accommodate, so that we did not confuse our meanings. The latter would be the kind of fatalism that focuses on options and events that are actual, and the

former, the compatibilist fatalism, would revolve around a kind of conditional fatalism that claims that no matter what we *would* have done, the outcome would have been the same.

Indeterminism is the denial of universal determinism and the affirmation that there are two or more possible futures given the current state of the world and the natural laws. Indeterminism allows for the possibility that there is more than one potential future state of affairs. In other words, the state of the world at  $t_1$  does not necessitate a particular state of affairs at all future times  $t_{1,2,3,4...}$  to  $T_n$ . It is not required for indeterminism that there be options at *every* time slice or that the number of options be infinite; it is only required that there be two or more options at some point in the future, defeating the inevitability of the future state of affairs of universal determinism and allowing for an open future, no matter how restrictive or expansive those options are. The source of this indeterminacy can be conceived as springing from a number of different places and, as we shall see, some may be more challenging than others to square with our requirements for certain kinds of moral responsibility.

### Classifying the Various Accounts of Moral Responsibility

There are a number of ways that accounts of moral responsibility can be categorized. We can differentiate, for instance, between those that are based on grounds for responsibility that are internal to the agent versus those that look to something outside of the agent to ground responsibility. We can also distinguish between accounts that are anchored to prior normative claims and those that have no such commitments. I will discuss both of these distinctions presently,

and will sort the accounts I discuss according to these classifications as I introduce each species of view.

### Agent-internal, Agent-external, and Included-agent Accounts

Accounts that base an agent's moral responsibility on the agent's possession, at least transiently, of some property or another that grounds this responsibility are agent-internal accounts. Accounts of responsibility that base the attribution of responsibility on something that is external to the agent are, unsurprisingly, agent-external accounts. Take, for example, a view that grounds moral responsibility on whether the agent has contra-causal control over her actions. This would be an agent-internal account, because the question of whether the person is morally responsible is answered by something about the agent herself, some feature of the agent in question, the feature of having regulative control. On the other hand, a view that focuses on whether there exist reactive attitudes directed at a given agent would be an agent-external account. There are, of course, accounts that include both agent-internal and agent-external criteria. I call these included-agent accounts, to differentiate them from the other two in which the agent is either central, or excluded.

### Autonomous and Anchored Accounts

Autonomous accounts of moral responsibility are able to support a definition of responsibility that does not rest on external commitments to normative claims. Anchored accounts, on the other hand, depend on additional normative commitments. These commitments may be concealed, but they are revealed on deeper examination of the entailments of the accounts.

To see the distinction more clearly, consider, for example, that someone could maintain that an individual is not morally responsible for a given action because he had no regulative control over the development of the events, but that we should, nevertheless, *claim* that he is responsible, for consequentialist reasons, because it will produce a deterrent effect in him and in others. This would still be an autonomous account of moral responsibility, because there are no external normative commitments that affect whether the agent *is* responsible. Despite there being a claim about whether one should *profess* that he is responsible, the normative claim is not prior to the determination of whether the agent *is* responsible.

Another might claim, instead, that not only should we assign responsibility to the above agent for the reason of deterrence or for other consequentialist reasons but that, as a result of this, the individual *is* morally responsible because an individual is morally responsible if it is consequentially justifiable to attribute responsibility to them. Such an account will be *anchored* to the normative view that fills in the criteria of being consequentially justified. For these anchored accounts the question of whether the person *is* morally responsible is answered by the question of whether it is justifiable *to assign* moral responsibility to that individual.

A proponent of an autonomous account can claim that an agent is not responsible, but that the ascription of responsibility is still justifiable. For the proponent of the anchored account, however, the question of whether an agent is morally responsible *depends on* the question of whether the ascription of moral

responsibility is normatively justified. It may seem as though the autonomous/anchored distinction overlaps with the agent-internal/external classification. However, the two groupings are distinct from one another. Although most agent-internal accounts are also autonomous accounts and most agent-external accounts are anchored accounts, the groups are not perfectly coextensive. A view can claim that there are things about the agent as well as criteria external to the agent that affect whether the agent is morally responsible, yet not be an anchored view, for instance.<sup>8</sup> This will become clearer as we look at the individual accounts. Unlike with the previous set of categories, there are no accounts that are a blend of the two poles. An account is anchored if it contains any criterion whatsoever that relies on prior normative commitments, even if it contains many other criteria that do not.

### Introduction to the Species of Views

Accounts in the debate on free will and moral responsibility are broken down into three broad categories: incompatibilism, compatibilism, and semi-compatibilism. Incompatibilists hold that determinism is incompatible with both freedom and moral responsibility. Among these views we find those that conclude that this undermines determinism, such as libertarianism, and those that conclude that it undermines freedom, such as hard determinism and hard-incompatibilism. Compatibilists claim that freedom and responsibility are compatible with determinism. There are a number of kinds of compatibilist accounts, including those that ground the existence of free will and responsibility

---

<sup>8</sup> Peter Strawson's view in "Freedom and Resentment" may be an example of this.

on freedom from external constraints, the ability to act in accordance with our deeper selves, the ability of our immediate desires to be subjugated by our reflective desires, evolved capabilities, reactive attitudes, and others. Finally, semi-compatibilists maintain that while we cannot have the kind of freedom desired by the incompatibilist, we can still have moral responsibility. Most of these views will be touched on individually below and further explained.

### Incompatibilism

Incompatibilists claim that freedom and responsibility are incompatible with determinism, requiring a kind of freedom that demands that the actions originate from within the agent without having been necessitated by anything else or by anything prior. Robert Kane calls this type of control “ultimate responsibility”, and he tells us that “to be *ultimately responsible* for an action, an agent must be responsible for anything that is a sufficient reason, cause, or motive for the action’s occurring” (Kane 121, emphasis is Kane’s). Most incompatibilists claim that in order to have ultimate responsibility, one must have what John Martin Fischer calls “regulative control”.<sup>9</sup> Regulative control is the sort that “involves genuine metaphysical access to alternative possibilities”. This is the kind of control that allows us to choose our actions from a set of alternatives (Fischer, *My Way* 6). Incompatibilists believe that we cannot conceive of a sort of moral responsibility that exists without this kind of incompatibilist freedom, and there are others, such as Immanuel Kant, who claim that we cannot even conceive of morality without it (Kane 33).

---

<sup>9</sup> Perhaps for some the direction of the requirement runs the way, as well.

According to the incompatibilist, determinism rules out both regulative control and ultimate responsibility. There are many arguments used in an attempt to support this claim. Among them, we find the Consequence Argument and many reformulated versions of it, the Impossibility Argument, the Manipulation Argument, and the Direct Argument, among others (Haji 207-208).<sup>10</sup> The Consequence Argument, put forward by Peter Van Inwagen, is the most commonly cited of these. Van Inwagen tells us that

“If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born; and neither is it up to us what the laws of nature are. Therefore the consequences of these things (including our own acts) are not up to us.” (Van Inwagen 16)

If our acts and their consequences are not “up to us”, the incompatibilist thinks that we cannot be morally responsible for either. Much debate between compatibilists and incompatibilists will center on just what it means for something to be “up to us” as well as whether the Consequence Argument is sound. Not all incompatibilists rely on the Consequence Argument, of course, but all incompatibilists accept some form of the premise, “If determinism is true, free will does not exist.”

### Libertarianism

There exists a group of incompatibilists who believe not only that free will and determinism are incompatible with one another, but also that free will is real, and therefore determinism is false. This view is called libertarianism. The

---

<sup>10</sup> For a detailed description of the various arguments, see Haji, 2009: “Incompatibilism’s Allure: Principal Arguments for Incompatibilism”.

libertarian argument uses the core incompatibilist premise and adds another premise affirming freedom. The libertarian argument is typically some variation of the following:

- 1) If determinism is true, free will does not exist.
- 2) Free will does exist.
- C) Determinism is false.

It would not be sufficient, of course, for the libertarian to demonstrate that determinism is false in order to prove the existence of freedom, and so the libertarian position rests on supporting the affirmation of free will, which may prove more difficult than it might first appear. The libertarian's most demanding challenge is that of providing an account of a type of freedom that can ground responsibility and that can exist without determinism, without resorting to metaphysically odd properties, powers, or entities. Indeterministic accounts of freedom and agency in which the indeterminism is found "in the wrong place" cannot "secure agential freedom" (Ekstrom 137n52), or at least not the kind of agential freedom required for responsibility. If, for example, there exists an indeterminism in the space between our choices (conscious or unconscious) and the initiation of our actions, it is hard to see how such interferences could undergird or enhance our freedom or responsibility.

### Hard Determinism

There are incompatibilists who affirm determinism, instead of affirming free will, as do the libertarians. The structure of their argument, which also includes the core incompatibilist premise, is as follows:

- 1) If determinism is true, free will does not exist.
- 2) Determinism is true.
- C) Free will does not exist.

Proponents of this type of view were labelled “hard determinists” by William James, in contrast with supporters of the pejoratively named “soft determinism” who endorsed the type of compatibilism that affirms determinism but “abhors harsh words and, repudiating fatality, necessity, and even predetermination, says that its real name is freedom” (117). Hard determinists claim that since the unfolding of all states and actions is determined, in the sense of necessitated, there is nothing to be added, removed, or changed by the agents. If the agents are not initiating or even contributing anything and their actions are merely the upshots of a sequence that was initiated long before their existence, it seems absurd to attribute blame to those agents for actions of which we disapprove or praise for those we admire. Therefore, hard determinists may be thought to have an additional fourth premise stating that without free will, there can be no moral responsibility and hence the conclusion that moral responsibility does not exist.

### Hard Incompatibilism

Like the hard determinist, the hard incompatibilist claims that free will and moral responsibility do not exist. Hard incompatibilists, however, claim not only that free will and moral responsibility are incompatible with determinism, but also that they are impossible whether or not determinism is true. The hard incompatibilist argument is something like the following:

- 1) If determinism is true, free will does not exist.
  - 2) If indeterminism is true, free will does not exist.
  - 3) Either determinism is true or indeterminism is true.
- C) Free will does not exist.

### Derk Pereboom's Hard Incompatibilism

Derk Pereboom's brand of hard incompatibilism is an example of this view. Pereboom claims that an individual cannot be morally responsible unless they are the original, initiating cause of their actions, or at least of the actions in question. Any agent lacking this kind of control also lacks moral responsibility (Pereboom 477). Pereboom compares actions that have deterministic causal histories with actions springing from third-party manipulation of an agent; for example, if an individual has an implant that would grant a third-party control of their behaviours. Although the "agent" would be performing the actions, they are not the original source of the causes of those actions and are, therefore, not responsible for them. Determinism, Pereboom thinks, is like the implanted chip, dictating the behaviour in which the individual will engage. He tells us also that actions produced indeterministically are similar to those produced by an implant that provides random instructions. In either of these cases, Pereboom explains, the agent does not initiate the causal chain of events that ultimately decides which action will be produced. In both deterministic and indeterministic causal chains, something other than the agent has control over the initiating causes of the agent's actions. This kind of control, he tells us, is the variety of control required for moral responsibility. In other words, neither determinism nor indeterminism make responsibility possible, leading to the conclusion that moral responsibility is impossible (Pereboom 477-478).

## Galen Strawson's Basic Argument

Galen Strawson has an argument for the impossibility of moral responsibility and the kind of freedom it would require. This argument is called the Basic Argument and is an appreciably powerful hard incompatibilist argument against moral responsibility. It appears, in one of its forms, as follows:

- “(1) It is undeniable that one is the way one is, initially, as a result of heredity and early experience, and it is undeniable that these are things for which one cannot be held to be in any way responsible (morally or otherwise).
- (2) One cannot at any later stage of life hope to accede to true moral responsibility for the way one is by trying to change the way one already is as a result of heredity and previous experience. For
- (3) both the particular way in which one is moved to try to change oneself, and the degree of one's success in one's attempt at change, will be determined by how one already is as a result of heredity and previous experience. And
- (4) any further changes that one can bring about only after one has brought about certain initial changes will in turn be determined, via the initial changes, by heredity and previous experience.
- (5) This may not be the whole story, for it may be that some changes in the way one is are traceable not to heredity and experience but to the influence of indeterministic or random factors. But it is absurd to suppose that indeterministic or random factors, for which one is *ex hypothesi* in no way responsible, can in themselves contribute in any way to one's being truly morally responsible for how one is.”

(Strawson, G. 7)

This argument is, I believe, the most compelling argument against the existence of the kind of moral responsibility usually discussed in the context of the free will debate. It is particularly powerful because it bypasses the question of freedom altogether, setting aside the potential for additional confusion introduced

by the debate over just what freedom is, and getting straight to the heart of the question of the possibility of moral responsibility. It is the fifth premise of this argument that is most vulnerable to opposition. There are some who will claim that there are indeterministic happenings that are not random but that are, instead, directed by an agent who has the power to influence the indeterministic course of events. Helen Steward's view, which will be discussed in the next chapter, is one such account.

### Classifying Incompatibilism

Incompatibilist accounts such as libertarianism, hard determinism, and hard incompatibilism are agent-internal and autonomous—accounts of moral responsibility. These views are agent-internal because the criteria for moral responsibility revolves around the presence or absence of a given feature within the agent; namely that of having the ability to be the efficient cause of her actions, what John Martin Fischer calls "regulative control", the control required in order to have genuine alternative possibilities or the ability to do otherwise. They are autonomous views because they do not rely on any prior normative commitments.

### Compatibilism

Compatibilists think determinism and freedom are compatible. The compatibilist is not claiming that we can possess the regulative control of the incompatibilist's variety of free will even if determinism is true. Instead, he claims that this kind of control is not required for free will and moral responsibility. The compatibilist believes the incompatibilist is mistaken when she demands this deeper control for what we call "free will" and that she is searching for something

that does not reflect our common understanding of this free will.<sup>11</sup> There is a wide assortment of compatibilist positions, each grounding their freedom and responsibility on different criteria. Soft compatibilists, as discussed above, hold the additional belief that determinism is true, though this is irrelevant to their views on moral responsibility, as the existence of determinism does not interfere with the compatibilist's variety of freedom.

### Classical Compatibilism

The classical compatibilist requires only that we be free from constraints, both internal and external, that interfere with our ability to do what we wish. To be free, in the classical compatibilist's sense, is

- "(1) to have the *power* or *ability* to do what we want or desire to do, which in turn entails
  - (2) an *absence of constraints* or impediments [...] preventing us from doing what we want"
- (Kane 13, emphasis is Kane's)

The stoics appear to endorse this brand of compatibilism. Classical compatibilists claim that although we are not able to choose our desires, we are free so long as we can obey them without constraints.

The "power or ability to do what we want" according to classical compatibilists is quite different from the power or ability to do otherwise required by the incompatibilist. Although both can be referred to by their proponents as "the ability to do otherwise", for the incompatibilist, this ability must have been present as a strong, metaphysically available option in the actual world. For compatibilists, on the other hand, the power to do otherwise means merely that I

---

<sup>11</sup> Incompatibilists have, of course, responded that it is the compatibilist who is failing to capture the common understanding of what it is to have free will.

would have had something like the ability to do otherwise *had I wanted to* or had *I attempted to*.

For a compatibilist like Daniel Dennett, one could have done otherwise if one would have done otherwise in one or more possible worlds that differ only slightly from the actual world, even if it is impossible to have done otherwise in the actual world (Dennett, *Freedom Evolves* 75-77). Dennett's brand of compatibilism adds to the traditional compatibilist view by requiring that the agent possess a set of competencies meant to ensure the required freedom from internal constraints as well as the ability to reason properly (rather than just the freedom from addictions and other compulsions). These include "flexibility of mind, general knowledge, social comprehension, and impulse control" (Dennett, *Freedom Evolves* 290-292).

### Second-order Desires and Reactive Attitudes

In addition to the classical compatibilist view, there are views in which freedom and responsibility depend on our second-order desires, our abilities to respond to reasons effectively, or our reactive attitudes. Harry Frankfurt tells us that while many animals want to do or refrain from doing this or that, humans have an additional ability to *want to want*, or to *want not to want* to do or refrain from doing this or that. He has called these latter types of desires "second-order desires". They are desires about our desires (Frankfurt 6-7). For Frankfurt, to be free is to have the ability to want what we want to want and to do what we want to want to do. It is to be able to make our wills and our behaviour abide by our second-order desires" (Frankfurt 15).

For Peter Strawson, we are free if we meet the conditions for the practice of being held responsible (Kane 107). More importantly, we are responsible if we are the targets of certain reactive attitudes, such as resentment, anger, forgiveness, and gratitude, or of a “demand for goodwill” (Strawson, P. 8, 10). Agents are not morally responsible if we have suspended our demand for goodwill and these reactive attitudes towards them in favour of an objective stance that abandons ordinary interpersonal attitudes towards the agent (Strawson, P. 13). None of these criteria rest on the truth or falsity of determinism, making both Peter Strawson’s reactive-attitudes-dependent account and Harry Frankfurt’s account based on second-order desires compatibilist accounts.

Although there are many other compatibilist accounts and, as a group, compatibilist positions are popular in the free will debate, I will not be spending additional time on them herein. William James declared the position of the soft determinist to be “a quagmire of evasion under which the real issue of fact has been entirely smothered” (James 117). While, unlike James, I may not be willing to dismiss compatibilism’s contribution entirely, I do not believe the answer for which we are searching lies in a compatibilist account of responsibility. While the freedom of the compatibilist genuinely may be one of the varieties of freedom “worth wanting” (Dennett, *Elbow Room* 139), it is not the variety in which we should be interested, as I believe compatibilists are in error when they attempt to ground any kind of merit- or deserts-based responsibility on it.<sup>12</sup>

---

<sup>12</sup> The task of supporting this claim, however, will have to be set aside for another time in order to keep the focus on the questions already at hand.

## Classifying Compatibilist Accounts

Classical compatibilism is an umbrella term that covers a set of included-agent views, including Dennett's. They rely on criteria internal to the agent, such as the agent's capabilities or competencies, but they also rely on features of the world external to the agent, on the constraints that stand between the agent and her goals. Frankfurt's compatibilist view based on second-order desires is an agent-internal account, as it depends exclusively on the agent's ability to square her behaviours with her second-order desires, on whether her actions match the behaviours in which she wants to want to engage. No features of the world external to the agent affect the agent's responsibility. Peter Strawson's reactive-attitudes compatibilism, however, is agent-external. The agent's moral responsibility depends on the reactive attitudes of others and their expectations of goodwill.

Many of the traditional classical compatibilism, second-order-desire compatibilism, and reactive attitude compatibilism accounts are autonomous accounts of moral responsibility and do not rely on normative claims in order to determine whether agents are responsible. Dennett's compatibilism, meanwhile, appears to have some normative entanglements, though it is not entirely clear whether they are prior to the determination of the agent as morally responsible. There appear to be times where he argues that someone is morally responsible *if* treating them as such will result in positive consequences. Other times, it appears as though he means to say that agents are morally responsible *and* that treating them as such will have positive consequences. Although I believe his

argument in *Freedom Evolves* relies on the first of these two options and is therefore an anchored view, thorough analysis of Dennett's position would be tangential to the goal of this project, and so the resolution of this question will have to be set aside.

### Semi-compatibilism

Semi-compatibilists such as John Martin Fischer, Mark Ravizza, and Ishtiyaque Haji claim that we cannot have the kind of contra-causal freedom that incompatibilists claim is required for moral responsibility if determinism is true, but they disagree with the incompatibilist claim that this type of freedom is in fact required for responsibility. In essence, they deny the compatibility of determinism and a *freedom to do otherwise*, but affirm the compatibility of determinism and responsibility.

### Fischer's Semicompatibilism

John Martin Fischer proposes a semi-compatibilist account as an answer to the moral responsibility problem. Fischer's view, like Pereboom's, requires that the agent play a specific role in the causal history of the action in question in order to be morally responsible. His view differs from Pereboom's, however, in that he believes it is "how the actual sequence unfolds — rather than the genuine metaphysical availability of alternative possibilities" that grounds responsibility (Fischer, "Précis" 236).

Fischer discusses two types of control: regulative control and guidance control. Regulative control, as discussed in the section on hard incompatibilism, is the kind required for the agent to have the genuine alternative possibility to do

otherwise (Fischer, "Précis" 240). Guidance control, on the other hand, has two sub-criteria: moderate reasons-responsiveness and mechanism ownership (Fischer, "Précis" 236-237).

If an agent can recognize that there are reasons in favour of (or against) an action, and can identify what those reasons are, she has what Fischer calls "reasons-recognition" (Fischer, "Précis" 237). Reasons-recognition is required for the more comprehensive reasons-responsiveness. An agent is reasons-responsive if she possesses reasons-recognition and can make, with regard to the reasons identified by that faculty, sufficiently good choices in connection to her reasons (Fischer, "Précis" 237-238).

The second of Fischer's criteria for guidance control, mechanism ownership, requires that the agent possess, in a sense, the mechanism that produces the behaviour within her. This mechanism ownership involves a certain degree of subjectivity. For an agent to have ownership of a mechanism, she must "[have] certain beliefs about [her] own agency and its effect in the world" (Fischer, "Précis" 237). Fischer tells us that the agent becomes responsible "in part at least by taking responsibility" (Fischer, "Précis" 237). In other words, the agent's responsibility depends, in part, on her belief in her own ownership of the mechanisms out of which the behaviour for which we are evaluating her responsibility springs.

Unlike Pereboom's account, which requires something like regulative control for moral responsibility, Fischer's requires only guidance control and so it requires neither genuine metaphysical control over one's actions, nor the

presence of alternative possibilities. Fischer's morally responsible agents need only to be able to recognize and act on their reasons and to feel that they have ownership over the mechanism from which that action is produced (Fischer, "Précis" 232).

Fischer believes his view addresses the question of whether moral responsibility is compatible with determinism (or, for that matter, indeterminism) by circumventing it. In this way, he believes he has constructed a view in which moral responsibility does not "hang on a thread" — the thread of the truth or falsity of determinism or indeterminism (Fischer, "Précis" 230, 240). Furthermore, he believes that this adequately addresses the "plausible kernel" that can be found within incompatibilist theories by showing that the regulative control on which they turn is not required. (Fischer, "Précis" 233-234).

Fischer's semicompatibilism, as he presents it, is an autonomous account with no normative commitments at the foundation of its argument. Additionally, it is an agent-internal view as the criteria for responsibility — reasons recognition, reasons responsiveness, and mechanism ownership — are all features of the agent, depending on nothing external to the agent for her responsibility.

Fischer's account is considered a semi-compatibilist account in the broader free will debate because it holds that free will is incompatible with determinism, as the former requires regulative control, but that both determinism and indeterminism are compatible with the guidance control and therefore, with moral responsibility. Although the view is classed as semi-compatibilism in the

free will debate, it is a compatibilist view of responsibility and will be treated as such herein.

Compatibilist accounts of responsibility (including Fischer's semicompatibilist account of responsibility and freedom) allow responsibility to be grounded by factors other than those within the agent's (regulative) control. Ultimately, I believe these positions either ground a non-deserts-based variety of responsibility — which they then mistakenly conflate with deserts-based responsibility or use as though it were deserts-based responsibility in order to justify further conclusions — or they unreasonably ground deserts-based responsibility on factors the agent could not, in the strongest sense, influence.

Of the accounts outlined above, it is the hard incompatibilist position, Galen Strawson's specifically, that I find the most compelling. It is, as mentioned above, vulnerable to an account that can show that some indeterministic happenings are not random, but are directed by agents with the power to affect the indeterministic sequence. Helen Steward's account is one that attempts to do precisely this. Her account, which will be discussed in the next chapter, does not merely pose a threat to hard incompatibilism, but takes aim at determinism and compatibilism as well.

### 3. Helen Steward's Libertarianism

#### Agency Incompatibilism

Helen Steward's variety of libertarianism defended in *A Metaphysics for Freedom* is one she calls "Agency Incompatibilism" (Steward 1). Steward claims that agency, a kind of agency attributed to non-human animals of a certain degree of complexity, is itself incompatible with determinism (Steward 1, 3). Agents, she tells us, are entities that settle through their actions (Steward 31). Seeing as everything in a deterministic universe is already settled, presumably by the state of things at the beginning of the deterministic chain in concert with the physical laws, agents could not exist in combination with universal determinism. She also claims that such entities as the agents she describes do, in fact, exist and therefore that determinism is false. Her argument resembles the following:

1. If universal determinism is true, everything is already settled.
2. If everything is already settled, there is nothing that settles anything.
3. Agents are entities that, among other things, settle certain things when they act.
4. Agents exist.
5. Therefore, universal determinism is false.

In order to further examine this argument, we must review Steward's conceptions of the notions central to her account. These include agency, action, and settling. The latter item, settling, is crucial for the understanding of the others.

#### Settling

Settling is the resolution of questions about what will happen, when, and how, that remain unresolved until the moment of settling (Steward 39). Settling

has occurred when an open question about what will happen becomes closed (Steward 41). In a deterministic universe or process, the states and events throughout the length of the deterministic chain are already settled. When I strike a cue ball in just the right manner to send it on a collision course with the eight ball, at just the right angle and speed to send the latter into the side pocket, then (assuming no one is near enough to the table to pick the balls up or to toss something into the eight ball's path, or otherwise disrupt its trajectory) the question of whether the eight ball will fall into the pocket is settled the moment the cue ball is struck by my cue.<sup>13</sup>

In a fully deterministic chain of events, the complete sequence is settled at the very beginning of the process, its details being fully fixed from that time on. In an indeterministic process, settling occurs as indeterministic events occur, closing questions about what will happen, when, how, or where, as those events happen. In a process in which determinism and indeterminism are combined, the deterministic chains that follow indeterministic events would be settled by the indeterministic events.<sup>14</sup>

## Agents

Steward defines an agent as:

"a creature that can move the whole or at least some parts of something we are inclined to think of as *its* body, that is the centre of some form of subjectivity, a creature that is something to which at least some rudimentary types of intentional state (e.g. trying, wanting, perceiving) may be properly attributed, and a

---

<sup>13</sup> Assuming it is not settled before this.

<sup>14</sup> Unless there were the persistent possibility of interference by other indeterministic events throughout the entire sequence of events.

creature that is a settler of matters concerning certain  
of the movements of its own body"  
(Steward 246; emphasis is Steward's).

For Steward, an agent is something that "things can be up to" and that plays an "irreducible role" in its own movement (Steward xii). Only creatures that embody the required level of complexity to play such a role can be agents (Steward 16, 32). As such, agents are "self-movers" and a self-mover "is able to *make itself (or part of itself) move*" (Steward 15, emphasis is Steward's).

Steward differentiates between entities that can "move by themselves" and entities that can "make themselves move" (Steward 15). There are entities for which it is not possible for us "to make the distinction between the creature and its body" and that, more importantly, are not complex enough to allow us to think of them as having a certain type of control over their own bodies in the way that complex creatures can (Steward 15). These are simpler entities that merely respond to external stimuli in predictable patterns (such as paramecium) or whose movements we would say are not fully up to them, even if that movement were randomized by internal processes and, therefore, not necessitated by anything (such as robots; Steward 17-18).<sup>15</sup> Steward believes it is top-down causation, made possible by a certain level of complexity, that allows self-moving creatures to have this type of input into their own actions.<sup>16</sup>

---

<sup>15</sup> Steward does not claim that no artificial life could ever make itself move, but that none can at this time (Steward 18n40).

<sup>16</sup> I will return to top-down causation below.

Although Steward believes top-down causation, and therefore settling, is present in inanimate objects, such as a whirlpool<sup>17</sup>, there are features that differentiate agents from other settlers (Steward 233-243). *Having* a body, possessing intentional states, and being at "the centre of some form of subjectivity", are the additional criteria that distinguish agents from other settlers. Nevertheless, settling of any kind, by animate agents or inanimate objects would disprove determinism.

Agency, according to Steward, is the capacity to settle how one's body will move — "the power of self-movement" (Steward 16, 164). It is "an *ongoing* capacity to stop or alter what [one] is doing in any of an almost infinite variety of ways, a capacity that is present throughout the whole of the period of activity, and not one confined to a certain point in a causal chain of interlinked events and states" (Steward 166, emphasis is Steward's). This capacity allows the agent to be "an initiator" because her actions have "no prior necessitating conditions" (Steward 246). As an initiator, the agent is "an ultimate arché, or origin" of her doings, because the chain of those actions and their results "cannot be traced back along lines of inevitability beyond her" (Steward 246). Agents' causings, Steward tells us, "are always at the same time causings by actions" (Steward 205). When an agent settles something, she settles it by acting or by refraining from acting.

---

<sup>17</sup> According to Steward, the whirlpool is an example of top-down causation because it is the "right kind of succession of molecular arrangements" that explains the whirlpool's movement, shape, and persistence. She claims that this succession of molecular arrangements that constitute the whirlpool is fully explained only by reference to the whole (the whirlpool), and not to the sum of the activity of its parts (the individual water molecules; Steward 241-242).

## Action

For Steward, actions are "events that are *causings* of bodily movements and changes by agents" (Steward 205; emphasis is my own). Steward also claims, however, that some actions are not events, but processes (Steward 45, 165). These causings of changes and bodily movements are to be distinguished from the *causes* of those changes and movements (Steward 45, 200, 205). Actions are the *movings* of one's body, not the resulting bodily movements (Steward 33). To borrow an example used by Steward throughout the book, my action is the raising of my arm, not the rising of that arm in response to my raising it. Additionally, actions are not the things that I do. 'What I do', Steward tells us, is a type of thing rather than a particular event. Actions, on the other hand, are my doings (Steward 152).

Steward tells us that these bodily movings are "causings of movements which are by us" (Steward 35). It is important to note that, although actions are *up to us*, Steward makes clear that actions do not require a decision or a choice. She cautions against claiming actions need mentalistic explanations, conscious or unconscious deliberation, or any corresponding mental states (Steward 173). In fact, decisions are, themselves, actions (Steward 155).

These bodily movings, these actions are not merely necessary consequences of prior events and conditions, they are active inputs into the world, interventions by agents that decide whether a certain movement or change occurs. They are *up to the agent* in the sense that they are the exercise of an agent's control over whether or not the changes occur (Steward 31, 78, 140,

174). They are the means by which agents settle what will occur, and how (Steward xi, 205-206). Through these actions agents settle, when acting or refraining from acting, whether, when, where, and how a particular change will occur. These are never necessitated prior to their actual coming about (Steward xi, 31, 38-39, 228). It is the fact that the matters remain unsettled until the moment the agent acts that makes them inconsistent with universal determinism. That agents are settlers of such matters is what extends this incompatibility to agency. If actions are agents' settlings of things left unsettled to that point, then actions and, by extension agents, are incompatible with determinism (Steward 43, 248).

### Refrainment

When an agent acts, she exercises a "power to make the body (or particular parts of the body) move" (Steward 32). This is a power that the agent need not have exercised when or how she does; the exercise of this power is never necessitated by antecedent events. It is a two-way power of refrainment, the power "to act or refrain from acting" (Steward 155). For an animal agent to be credited with settling, the animal must have this power to refrain from the activity in which it engages, at the time at which it engages in it. It must have the power either to do something else or to do nothing at all (Steward 126). No movement or change can be regarded as settled by an animal or as that animal's action, according to Steward, if the animal did not have the power to refrain from that movement or change (Steward 126).

## Joe's Decision

Steward provides us with a thought experiment about a man name Joe who has the opportunity to move in with his girlfriend. This thought experiment illustrates the role the agent plays in settling his actions as well as what kinds of things are up to the agent, and how. It will serve as an example throughout the discussion of Steward's account. She tells us that Joe has both reasons and desires that favour moving in with his girlfriend, and that he is aware of them (Steward 128-174). Steward thinks that any variety of freedom that would allow Joe to choose anything that was in conflict with his reasons and desires would fail to be a kind of freedom the libertarian should want for Joe.<sup>18</sup> Steward tells us that the type of freedom the libertarian should be trying to explain and defend is the freedom, not to do something different, as Joe would, for example, if he chose not to move in with his girlfriend, but the freedom to do the type of thing that he will do differently, or to refrain from doing anything altogether. She explains that Joe cannot decide not to move in with his girlfriend, but can delay his decision by minutes or days, perhaps even long enough for his girlfriend to become frustrated and withdraw the offer (Steward 168).<sup>19</sup>

Steward claims that if there were two identical worlds with identical histories and identical Joes, the Joe in the first world could make the decision in

---

<sup>18</sup> Or for us.

<sup>19</sup> I question whether Joe could, in actuality, delay his decision long enough to frustrate his girlfriend and cause her to rescind the offer. It strikes me as unlikely, especially if he senses her growing frustration and all his conative states favour the move. This may not be important for questions at the heart of the debate about indeterminism and the agent. Though, it does lead one to wonder about how much responsibility we can ground if the constraints to our behaviour placed upon us by our conative states might be strong enough to determine that certain types of activities are guaranteed, even though the token may be different.

one way, and the Joe in the second world could make the decision differently. For example, the Joe in the second world could make the decision a few minutes later than the Joe in the first world. The difference between the behaviours observed in these two worlds would, she believes, be up to the Joes, each of them settling how they will act and therefore settling how the future in that world will unfold. According to determinists, the two Joes — I will refer to them as Joe and Joe\* — would behave in precisely the same way in both worlds, given that they are identical, subject to the same physical laws, and share the same history and current state of events. The determinist tells us that both Joes will make the same decision, will make the decision at the same time, and will make the decision in the same way. Steward disagrees, saying that although they cannot make utterly different decisions, the one deciding to move in with his girlfriend and the other deciding not to, they can make those decisions differently by doing something like making them at different times, or refraining from making a decision altogether for the time being.

In this thought experiment, there can be no differences between the two Joes. Since the differences in Joe's and Joe\*'s behaviours are not accounted for by there being any differences between them, a natural response might be to wonder what, other than the Joes' features, accounts for this difference in their behaviour. It seems perhaps natural to ask whether, if the difference in behaviour is not accounted for by differences between the Joes themselves, not accounted for by anything about the way either or both of them are, how it could be that the difference is settled by either or both of them, at all. We might think that if the

difference is settled by something other than the way they are, by something *about them*, then it is settled by something other than them. Of course, determinists and hard incompatibilists alike are likely to say that if the difference were settled by something about them, it would really be settled by whatever it is that settled that they would be as they are. Yet, if the differences are settled by something else, whether chance or some other external indeterministic process, it seems that it would not be any more "up to them". Both of these options eliminate the agent's settling from the picture, whether indeterminism is true or false.

### Ontological Pluralism about Causes

Steward tells us, however, that agents' properties "do not dictate that they will act, in given circumstances, in just one possible way" (Steward 131). She argues that properties are not the types of thing that are capable of *doing* anything, as they are not causal agents. Steward proposes a pluralism about the ontology of causes. She claims there are three types of causes: movers, matterers, and makers-happen. Movers are what we think of as substances. They are the rocks, the windows, and the matches. They are also things such as molecules, atoms, and perhaps fields. They are spatiotemporal particulars. Makers-happen are triggering events. They are the throwing of a rock, the kicking of a ball, or the striking of a match. These are also particulars. Finally, matterers are facts. They are the facts that would follow the word "because" in a statement. For example, the rock broke the window because it was heavy. The match would not light because it was wet (Steward 212-214). Steward holds that only movers

can *do* things, but properties are matterers (Steward 213n36). Much like other matterers, such as other facts or, I presume, the physical laws, properties may only constrain the agent's actions, rather than necessitating them. For Steward, it is not our properties that settle the details of our actions. It is not our properties or any facts *about* Joe or Joe\* that settles what they will do, nor is it chance; it is Joe and Joe\*, themselves, who settle the specific details of their actions.

### Top-down Causation

On Steward's account, this ability to settle the specifics of our actions is due to top-down causation. According to proponents of top-down causation, the behaviour of an aggregated whole can be greater than what can be explained as, and predicted by merely the sum of the behaviour of its parts. At a certain level of complexity, the whole itself can direct changes in its component parts, changes that are more than, or different from, the changes that would occur as a result of the effects of the component parts on one another in a bottom-up arrangement. The parts and their arrangement at any given time do not necessitate the arrangement of the parts at a future time, though they constrain those future arrangements, to some extent.<sup>20</sup>

This results in a supervenient whole for which the structure at  $t_2$  is not necessitated by the structure at  $t_1$  or by the properties, relations, or movements of its constituent parts at  $t_1$ . The actions of this supervenient whole, then, are not necessitated by its constituent, subvening parts when the whole is complex

---

<sup>20</sup> Although I am skeptical of the tenability of top-down causation, I will not be arguing against it herein. I will, instead, be granting the assumption that it is plausible and will leave the more complex discussion of whether top-down causation accounts are workable to seasoned metaphysicians.

enough to induce top-down causation. The agent's ability to settle is allowed by the fact that the laws of physics merely constrain, rather than necessitate, and complex wholes can give rise to top-down causation, allowing those wholes to direct the coordination of the parts (Steward 239-244). Causation, then, can operate along the level of the supervenient wholes, at the macro-level, rather than along the micro-level of the supervenience base. Steward believes that in a picture such as this one, there is room for the animal agent, as a complex whole, to be the director of its own actions, removing the animal's parts from that role. As a result, the processes that cause the agent's creation or settle its properties at any given time do not also settle that agent's actions at any future time. The agent's movements and changes are her own and are settled by her because the details of those movements and changes are not necessitated by features of the supervenience base. They are not necessitated by any prior states or events, either internal or external to the agent, nor are they necessitated by the physical laws, though all of the former may constrain them. Steward believes that her account of an agency that is braced up by top-down causation places the agent at the centre of the initiation of her actions, at the helm of the causal vehicle, so to speak.

#### Problems with Steward's Account: A Unique Kind of Settling?

It is perhaps unclear up to this point, however, how it is that an agent's settling is different from the other causes and constraints which cause without necessitating. What differentiates the agent such that she is not merely an additional non-necessitating ingredient in the causal admixture that gives rise to

her movements and changes? Steward claims that the agent also settles without necessitating (Steward 150). Yet, when she does so, she appears to be referring to the agent being unable to necessitate the effects of her actions, as opposed to her actions themselves. She tells us the following:

"When I determine that something will happen, that is, I do not necessarily render it inevitable that it will. I do not necessarily necessitate anything (though it is true that it is good for me if my action is merely the last remaining part of a sufficient condition for the result at which I am aiming). But as I act, I bring a state of affairs into being that need not have been brought into being, for my action itself is something that need not have happened: it is not itself necessitated by any prior conditions. And this is what enables me to settle or determine what will occur"

(Steward 151).

In this excerpt, it seems that Steward is referring to something different from the action itself, she is referring, instead, to the outcome of the action, but what of the "action" itself? What of the movements and changes made in the agent's own body? The agent may indeed be a kind of cause, a mover, but so are rocks and matches. What is supposed to be interesting about the agent is that it is the agent that settles at least some of the details of some of her movements and changes, whether or not those movements and changes necessitate any given outcome.

The agent qua mover and event qua matterer are both causes of the outcome. The settling by the agent of her own actions, however, is what is of interest. If the agent's actions can originate from within her, unnecessitated, there can be room for merited praise and blame for her actions, even should the agent be unable to guarantee a given outcome of a given action. If, for example, I point

a gun at you while you are standing across a street from me and fire, I am, it would seem, equally responsible for pulling the trigger of the gun whether my bullet hits you in the knee or a bus pulls between us at precisely the right moment to cause my bullet to ricochet, giving you time to flee before I can get another shot off.

It might seem odd to say that I am not responsible for shooting you, only for pulling a trigger with the intent to shoot you, but consider another situation in which I am at a shooting range, engaging in target practice. This time, the altering event in question is you falling in front of the target just as I fire (perhaps from a skylight as you attempt to crawl across the roof to break into the range, not knowing that it is open on weekends), resulting in you being hit by my bullet. It is unlikely that I would be considered blameworthy for shooting you. I would be considered responsible, perhaps, for shooting my gun with the intent to hit the target (assuming that I am in fact settling the movings of my fingers that result in my pulling the trigger), as that was my intended activity (and let us also say the activity I should have been able to assume, without carelessness, was going to happen), but not for shooting you.<sup>21</sup> Although 'what I do' (i.e. the type of activity: either shooting you in the knee or shooting the side of a bus; either shooting a target or shooting you as you fall in front of it), is altered by the events that followed my action, my action itself (the finger movements sufficient to activate the weapon's trigger) remains the same and purportedly remains that which I am settling. In these scenarios, however, 'what I do' is no more settled by me than it

---

<sup>21</sup> This, of course, highlights the importance of conative and cognitive states as well as epistemic access to potential outcomes in the assigning of merit-based moral responsibility, but this discussion will need to be saved for a later time.

is by the accompanying event-causes. If we now suspend the assumption, granted above, that I alone was settling *the action* (whether I moved my trigger finger) and examine it more closely, can we retain a kind of settling that is by the agent in such a way as to allow for responsibility?

In Steward's picture, the agent's action, but not what the agent does, is settled by the agent and not by a group of indeterminate factors that together result in a settling; the agent is not merely "a crucial component of the necessitating mix" (Steward 150). It will not help the agent any more, however, if she is merely a crucial component of a non-necessitating mix. A clear account needs to be given about why the agent is dubbed a "settler" while the other indeterminate, non-necessitating causes that lead to the occurrence of the action are excluded from this category. The answer appears to be that the actions are not *settled* until just exactly what will happen is fully fixed and, given Steward's claim that agents have a two-way power of refrainment that allows them to alter or refrain from performing a given action before and throughout its execution, what will happen is never settled until the agent acts. The agent's movements and changes are never fixed until they have been performed.

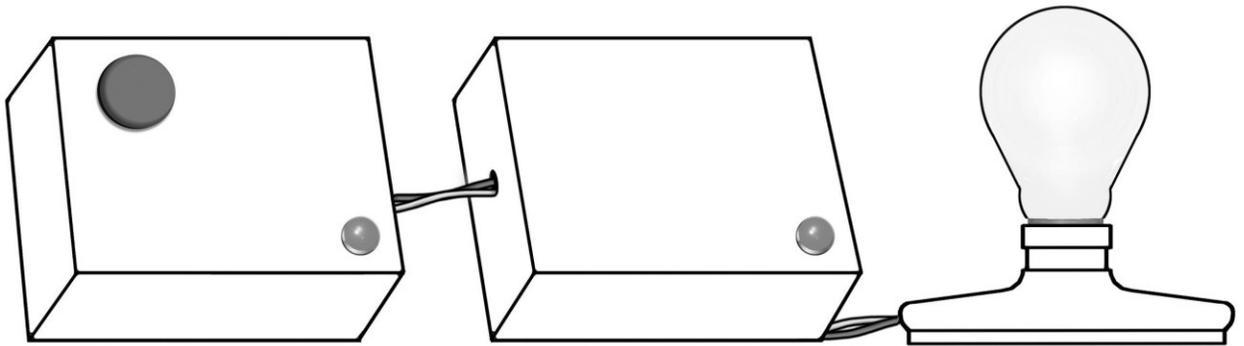
While it may be that the agent is the last indeterministic element in a chain that involves movements and changes of that agent's body, and that it offers the final contribution before a movement or change is settled, it is questionable that it should retain sole ownership of the status of "Settler" for that movement or change. The agent's contributions, when the agent is "acting" come last because they must. Since the changes happen to *them*, if there are indeterministic

happenings within the physical boundaries of what we are calling “the agent”, after the agent is influenced by the prior events, we are expected to say that the agent has settled, and therefore that the agent has acted. If there are no such indeterministic happenings within the agent, we are to say that the agent was deterministically influenced, and so the agent’s changes are not actions. Is this enough to provide a meaningful (or even an interesting) kind of settling? If the agent is not an entity whose movements are deterministically necessitated by prior circumstances, then it will be contributing something to the settling of those movements. But is the fact that that contribution comes last enough to give it a special status? If Steward’s agent is merely performing a kind of settling of this type, it seems not to be adequate for the type of moral responsibility found in the attribution of praise and blame.

Event-causes, Steward tells us, must “[operate] through my agency, not in competition with it” (Steward 219). As a result, there is room for me to *inject* settling into the causal chain, making it such that the event-cause does in fact cause an action, in a sense, but does not necessitate it. It is puzzling how the agent avoids taking this same role of causing without necessitating. It is unclear how it is that physical laws and event-causes are said to cause without necessitating, yet do not contribute to settling, while agents are not rendered equally impotent as far as settling is concerned. Why is the animal agent given a kind of priority and labelled as the ultimate “Settler” over the other indeterministic, non-necessitating causes?

### More Problems in Steward's Account: "Settled When" vs. "Settled By"

Let's assume there are two random number generators (RNGs) connected in a series (see figure 3.1). There is a button that initiates the operation of the first RNG and a deterministic process between the first and the second RNG, and another between the second RNG and a circuit that switches on a light. If the result of the first RNG is an odd number, the second RNG will be triggered to select a number as well. If that RNG produces an odd number also, completion of the circuit will be triggered and the light will turn on.



(Figure 3.1)

Like prior non-necessitating event-causes, the first RNG's results will constrain but not necessitate either the outcome for the light or the activity of the second RNG (analogous to the agent). If the first RNG produces an even number, the light will not turn on, nor will the second RNG produce a result. If the first RNG produces an odd number, the outcome will not be necessitated, nor will the result of the second RNG. Since the process between the second RNG and the circuit and its light are deterministic, the question of whether the light will turn on is settled at the time the second RNG's result is fixed. Still, despite it being settled at the time of the second RNG's activity, it seems quite odd to say that it

is the second RNG, exclusively, that settles that the light will turn on after the button is pressed. The first RNG plays at least an equivalent role in the resolution of the question.

One might respond that the light coming on is equivalent to the *outcome* of one's activities, which the agent cannot settle, rather than to the settling of actions themselves. However, seeing as the light being lit up is a purely deterministic upshot, we can back up to the point in the chain prior to this without losing anything of significance from the example.<sup>22</sup> In other words, we can say that it seems odd to claim that the selection of an odd number by the second RNG is exclusively settled by the second RNG; it still seems that the first RNG plays at least an equal role in the settling of the matter of whether the second RNG will produce an odd number. The first RNG cannot fully settle that the second RNG will produce an odd number, but it does seem that it plays a significant role in the settling of that matter nonetheless. Additionally, it does not merely play a limited explanatory role discussed by Steward. It plays a significant role in the settling of the question of what it is that the second RNG will do, whether it will select an odd number. While the question of whether the light will turn on is settled *when* the second RNG produces a number, it is not settled *by* the second RNG exclusively. The kind of settling about which we would be talking, if we were merely saying that the second RNG would settle which of the numbers was selected once it was prompted to select a number, that it would be adding the final indeterministic ingredient, would be a vacuous kind of settling as

---

<sup>22</sup> Why include the light to begin with? It is intended to help broaden the way we look at the interconnectivity of even indeterministic processes, rather than seeing each of them as small, independent islands of settling.

far as responsibility is concerned. Responsibility will require more than merely being the last step in a series of perhaps equally influential happenings. This does not undermine Steward's argument that if there is any settling at all, the determinism is shown to be false, but it does make us question labeling the argument as one that can support a libertarianism that undergirds responsibility.

### Physical Realizations

Steward explains that "[action], after all, is not magic; it needs a physical realization if it is to create physical effects such as bodily movements" (Steward 227). It may be the animal's physical realization that gives it special importance over the event-causes, physical laws, and prior conditions that are not considered settlers of the actions in question. Steward also tells us that animal agents are of a level of extreme complexity, such that their changes and movements originate from more than simply the collection of changes and movements of their constituent parts. The parts, she claims, are not complex enough to free themselves from either deterministic causation or the mere influence of chance events (Steward 220-222). Perhaps, then, it is the combination of both the top-down-causation-instantiating complexity and the physical manifestation that separate the animal agent from the other non-physical causes. This, however, under-explains why the animal agent is differentiated from other complex physical structures; structures complex enough

to allow top-down causation. Why, for example, is the whirlpool given different treatment in discussions about causation than the agent?<sup>23</sup>

The whirlpool has a physical manifestation and Steward believes it is complex enough to be an example of something that can enkindle top-down causation (Steward 241-244). Are the movements and changes of a whirlpool to be considered actions? If not, is this merely due to the lack of the other criterion, laid out by Steward, of having intentional states, being at the centre of a kind of subjectivity, and having a body (Steward 246)? The criteria of having a body does not seem robust enough to make the difference between something like a horse and a whirlpool. Both the horse and the whirlpool's physical manifestations seem equally important to the causal story. It is not the body itself that is the important piece to this set of criteria, but the ability to make it move. This, however, is permitted by the top-down causation present in both of these cases, and so cannot be that which makes the difference.

Having intentional states and being at the centre of a kind of subjectivity may set the two apart for the purposes of ethical questions and perhaps for questions of responsibility, but I fail to see how they could make a difference in the metaphysical categorization of the types of causes each of these is. It seems reasonable enough to state that an agent is a top-down causation object or

---

<sup>23</sup> A wheel rolling downhill, which she also proposes may demonstrate top-down causation, would be a better example here, as its boundaries are more clearly defined in comparison to the whirlpool, and so would make a more natural comparison to other downward-causation-instantiating objects. Much more is said in the book about the whirlpool, however, and so it is a more natural point of comparison (Steward 233). Nevertheless, I would like to preempt the potential response that a whirlpool may be a phenomenon that is not a mover by pointing out that Steward does appear to support the claim that there are other, non-agent movers that instantiate downward causation, and I invite anyone who is blocked by features of the whirlpool example to replace it with the example of the wheel rolling downhill.

phenomenon, with the additional mental states and subjectivity, but the picture of agents that Steward is attempting to paint appears to be one that is more than simply something that resembles whirlpools with consciousness. Furthermore, we would be remiss if we failed to question why the level of the animal holds the special privilege of being labelled a settler in a system that includes top-down causation. If the whirlpool is complex enough to instantiate top-down causation, consistency may require us to continue looking to higher macro-levels to see if they do not act upon us the way Steward thinks we can act upon our parts.

The universe<sup>24</sup> exists as a physical realization and is certainly complex enough, if an animal or human is, to meet the complexity requirements to induce top-down causation adequately. Why would we not look further upwards to the galaxy or universe as the appropriate level in order to explain top-down causation? Would the universe not be the settler, then, of the movements and changes of its constituent parts? The movements of each animal and its constituent parts, being no more than constituent parts of the whole that is the universe, would be determined, top-down, by that whole. It would be the universe, then, that settles those movements and changes we are currently attributing to the animal in the same way as Steward's picture envisioned the animal settling the movements and changes among its own constituent parts. In this case, it seems there would be no role as a settler left to be played by the animal "agent", no reason to grant a privileged level of settling to any animal. Without such a reason, we cannot crown the animal as the level of the initiation of action unless

---

<sup>24</sup> Or the set of multiverses, if you prefer.

done arbitrarily. Through a complete top-down picture the agent is swept away, again, from a position of causal power. Not, this time, by the movements and changes of its parts, but by the movements and changes of the whole (or wholes) of which it is merely a part.

Someone could respond to this challenge by saying that it may be the case that the direction of causal efficacy flows from the agent outward, both upward and downward, and that the two (or more) levels clash. Which level wins is then resolved by a random process of some sort, and the settling of the matter is taken from both the agent and the universe and handed to the random process that settles which gets its way, so to speak. If the process is not random, however, if there is a predictable and necessitating process of some sort that makes it such that one or the other of the levels always settles, then the one that always triumphs is the one that is the true settler. For example, if the agent is always the settler in these clash situations it will eliminate the possibility of the universe's top-down causation powers over the animal, putting us back in the position of having to explain why the agent instantiates downward (and upward) causation, but the universe does not.

Someone could also reply that the agent's actions are constrained, but not settled, from the top, but that the animal agent still settles, top-down, what will happen next. This, of course, leads us to wonder if the animal is truly the last of the things to be able to do so. Is it not possible, then, that the animal only constrains other indeterministic processes at lower levels of supervenience, leaving the final settling to something else? Are we simply going to claim that the

animal agent is the least complex of the supervenience levels that are complex enough to instantiate top-down causation?

Another could claim that both levels can constrain but not settle each other, and this seems perfectly plausible when we are thinking of a single universe and a single agent. If, however, there are many, many agents settling — easily over a hundred billion if animals are settlers — then the web of constraints will be extremely limiting. The extent of the limitations would place such severe restrictions on each component that, if it is truly feasible that there is any indeterminacy left at all in such a web, it would be found well below a level that would generate the opportunity for settling anything that could possibly affect a behavioural level, the level at which moral responsibility comes into play.

#### Two-way Power of Refrainment: An Inadequate Defence

It may be the two-way power of refrainment that differentiates agents from the other settlers. Steward tells us that agents have the

"constant and ongoing capacity throughout the period of the action to do such things as: cease from an activity in which she is currently engaged, alter the speed of some movement in which she is currently engaged, change the direction of a movement, begin new types of movement, either instead of, or as well as, those already being engaged in; etc."

(Steward 164).

Steward claims animals possess these two-way powers of refrainment and alteration whether or not we are currently exercising them (Steward 46). Perhaps only agents are able to wield this two-way power, differentiating them from other top-down-causation-realizing objects. These powers, then, would have to differ from any powers possessed by the wheel rolling downhill, by the whirlpool, or by

the universe. But of what would such a power consist? Would such a power be enough to override any top-down causation by the universe? And if so, is it possible to provide a naturalistic explanation of the source of this power? We are at risk of creating odd and perhaps inexplicable metaphysical phenomena if we are unable to give a proper account of why this power exists among animal agents but not whirlpools. Yet, if the whirlpool also possesses this two-way power of refrainment, we are again backed into the corner of needing to explain the difference between agents and whirlpools and universes. The plausibility of Steward's account may rest on this explanation. Merely assuming agents can possess such a power uniquely may beg the question in favour of the libertarian account. More seriously, it may risk a flirtation with dualism. We need a full account of why this power differs from any power possessed by a whirlpool; one that explains from where it springs and how it occurs.

Even if we are provided with an adequate account of this power — where it resides, what it is, why it is only agents who possess it — an explanation will need to be given as to why Joe and Joe\* use the power differently. If the power is what allows Joe and Joe\* to behave in different ways, what is it that allows Joe and Joe\* to use the power differently? Why do they not use it identically, seeing as they are identical and possess identical powers? The two-way power of refrainment cannot explain itself or its own use. How, then, can we explain the different use of this power by Joe and Joe\* in a way that does not rely merely on chance? If the use of this power explains the ability to use the power, we find ourselves in a position where we are supporting a kind of bootstrapping. We risk

positing an odd and metaphysically irregular power. It may appear as though we have eliminated the self-creating, bootstrapping features of past libertarian accounts of agents, but it has simply been shuffled deeper, into the idea of the two-way power.

To further assess this problem, consider that at  $t_1$  Joe and Joe\* are identical, are in identical conditions, share their powers identically (including the two-way power of refrainment), and have used their two-way powers identically to this point. What can explain why each Joe uses his power differently? When an animal agent acts, there will be top-down changes in the supervening whole and its subvening parts between  $t_1$  and  $t_2$ . The animal's actions are those changes. Each time slice tells us what the animal has done (or is still doing) when compared to the previous time slice. It also tells us how the animal is, or how it is *being*, the shape the animal has taken. The top level of supervenience is not a level that represents action, but one that represents the shape of a supervening whole. It is this whole, when that whole is an agent, that will act. Steward concedes that supervenience on the physical requires "that no two worlds that were identical in all basic physical respects at any given time  $t$  could differ in any supervenient respect at that same time  $t$ ". She adds,

"Even if [...] there may not be a change in any supervenient property without a change in the 'base' properties on which it supervenes, we have not yet arrived at the idea that any given supervenient change must depend for its explanation upon the subvening one that [...] must always accompany it. The direction of explanation might (in some instances) rather be the other way about."

(Steward 120)

Steward has given us an account on which the supervening whole is necessitated by neither the prior supervening whole, nor by the prior supervenience base. Even granting that the direction of explanation might flow from the top down, the supervening and subvening levels change concurrently. Additionally, the current bearing and speed of those movements and changes are features of the constituent parts of this whole. As a result, in a scenario in which Joe and Joe\* are identical at  $t_1$ , they must also be "acting" identically, at least at that particular time slice. When the changes begin to differ from one another the Joes will immediately *be* different from one another. In order for them to begin behaving differently at  $t_2$ , they will need to be different at the first point of divergence in their behaviour. The differences at the first sign of change cannot be accounted for by any differences between them at  $t_1$  as they were identical at that time. This is beside the point, in any event, as claiming that the differences between the Joes at  $t_2$  were fully accounted for (in that they were necessitated) by the Joes at  $t_1$  would beg the question against indeterminism. By  $t_2$ , by the time they are acting differently, there is already dissimilarity between them; just as soon as they are acting differently, they are already different. They are no longer identical Joes and since we cannot account for this divergence by differences between them at  $t_1$  they must be accounted for by something else. Steward tells us that Joe and Joe\* possess a two-way power of refrainment. These changes, however, cannot be attributed to the exercise of a two way power as the Joes could not exercise their powers differently without a corresponding difference in their subvening physical bases. By the time the Joes have used, or begun using,

a two-way power of refrainment differently, there will already be differences between them and those changes will either have to be attributed to that power itself or to something else. It cannot be something else, however, as that power *is* that which allows them to change in different ways from one another.

At the point at which they begin exercising their power differently, there must already be changes that allowed them to do so. This means that their powers must already have been exercised, in order for them to be doing something — exercising their powers — differently. At the precise moment in which the Joes begin exercising their power differently, have they not become, at that same moment, different? What is the precise moment at which the Joes begin exercising their power differently? Are they not, as soon as the different exercises of these powers begin, already different?

The way Steward describes the two-way power, we get an image of two Joes, identical to one another, and we think of Joe directing his power such that he makes a decision and Joe\* directing his power such that he waits a moment, he refrains. It seems as though there is the power and there is its exercise, and all changes, all differences in the Joes, happen at a time after each exercises his power. The exercises of the power, however, given the account of supervenience used by Steward, quoted above, would be reflected in concurrent changes, changes that happen *as the purported power first begins to be exercised*. The power being exercised, then, can be seen not as initiating the first change, but as the first change itself.



not supervene on a subvening physical base. Regardless of the direction of dependence, the problem remains. The power appears to need to initiate its own initiation. The exercise of the power finds no temporal space in which to happen as it must be its own cause and result.

A third response to this, one I think Steward might be more likely to accept, is to claim that the two-way power is an emergent power of some kind, one that does not require supervenience yet can exist within a physicalist paradigm.<sup>25</sup> Like the discussion of top-down causation, however, the discussion of the plausibility of this kind of emergence is beyond the scope of this project. Furthermore, if the power explains the movements and changes in the agent, what explains why this is *the agent's* power? Why would this not be considered to be a power belonging to the totality of indeterministic reality, engendering changes here and there, whether inside or outside of animal agents? And if it can be conceived of in this manner, Steward's account may be vulnerable to Mele's Challenge from Chance.<sup>26</sup>

### Grounding Responsibility

In addition to the question of whether top-down causation is plausible and whether we can make sense of this two-way power of refrainment without an odd kind of bootstrapping, there is also the question of whether the agent is merely one of many non-necessitating causes. Though the above challenges may be met and the agent shown to be able to settle, it seems as though the best kind of

---

<sup>25</sup> For example, consider the kind of emergence proposed by Timothy O'Connor in the last decade.

<sup>26</sup> For Steward's discussion of Mele's Challenge from Chance, see Chapters 6 and 7 of Steward, 2012.

settling that Steward's account can attribute to the agent (if the agent is to be the sole settler of its movements and changes) is a settling in the sense of closing some previously open question(s) about minute details of how something may happen. Any more significant settling, any settling robust enough to ground responsibility, would require a wider scope and the inclusion of additional settlers. This is not Steward's project and so this final challenge about responsibility is not aimed directly at her agency incompatibilism itself, but it does show how this conception of action and agency could not be a threat to the hard incompatibilist about praise and blame, nor could it ground responsibility in the way I suspect many libertarians hope.

## 4. Responsibility in the Sense of Onus

### Responsibility as a Divisible Concept

In this section, I present an alternative to the views discussed so far. In many of the above accounts, moral responsibility is treated as one large concept, indivisible and threatened in its entirety by various elements discussed in the free will debate, determinism specifically. In actuality, “moral responsibility” is used in different senses in both philosophical and everyday discourse. It is crucial to differentiate between these senses and, when we do, we see that one of them is compatible with determinism and indeterminism.

It has been argued, skilfully and successfully, I believe, by individuals such as Derk Pereboom and Galen Strawson, that there can be no moral responsibility of the kind required to ground merit-based praise and blame, regardless of the deterministic or indeterministic nature of the universe. Although the sense of responsibility I believe they have in mind, responsibility qua causal accountability, is shown to be illusory, there is another sense of responsibility, responsibility qua onus, that I claim survives regardless of determinism, and that can enter the fray of the free will debate and come out wholly unharmed.

What I am calling “causal accountability” is intended to be the variety of accountability we mean when we say, “He is responsible for failing to save the drowning baby from the puddle, and so should be blamed.” If one is persuaded, as I believe we should be, by the arguments put forward by Pereboom and Galen Strawson, then in order to have this kind of responsibility, we must have genuine control over our actions and their causes (Pereboom 477). It is this kind of

responsibility that requires Fischer's regulative control and that grounds praise and blame in an incompatibilist, non-instrumental way. If determinism is true, we do not have this regulative control. As a result, if determinism is true, we cannot have this responsibility qua causal accountability.<sup>27</sup> Additionally, if the hard incompatibilists are correct, as I believe them to be, there is no way at all for us to have this variety of responsibility, even if determinism is false.

We use "responsibility" in other ways, however. If we say, "She has a responsibility to ensure her articles of clothing are purchased from manufacturers who use ethical manufacturing practices", we are using a different sense of responsibility. We are indicating that there is a moral requirement for her to monitor her purchases; that there is something that she *ought* to do. This sense of responsibility is the sense to which I refer as "responsibility qua onus". I have a genuine responsibility qua onus when I am the target of an appropriate, future-regarding "ought" belief or claim.

### What Ought I to Do?

Before I discuss what is required for an ought belief to be appropriate, I would like to consider the following thought experiment. Let us say that I am on the pier at my lake house and I see someone fall out of a boat that then drifts out of his reach. He is not wearing a life-preserver, and he appears not to be keeping his head above water consistently. He is quite a far distance from me, and so I have to strain to see him. As I look more closely, it looks as though he is at a great risk of drowning. I own a motor boat that is only a few steps away from me

---

<sup>27</sup> I will refer to this as simply "responsibility qua accountability", rather than "responsibility qua causal accountability" for the remainder of this work.

with which I could reach him in a short enough time to help him. I form the belief that I have the responsibility to get in the boat and head out to save him. Shortly after I unhitch the boat and begin making my way towards him, the motor stops cold and will not start again. I have a paddle but I am very aware that the distance between us will mean that he will not survive the time it takes me to reach him. As I repeatedly try to start the motor, it becomes clear to me that I cannot, and even if I could, reaching him in time is now impossible. Now that I know with certainty that I cannot save him, that it is impossible for me to save him, would I say that I still have a responsibility to do so?

If “ought” implies “can”, it seems, correctly, I believe, that I cannot say that I ought to save him, now that I know that I cannot. I would be mistaken if I were to continue to hold myself responsible for doing that which I cannot. Now that I know that it is impossible, I abandon my belief that I ought to save him and, with it, the idea that I have a responsibility to save him. I have neither the responsibility qua accountability — I am not blameworthy for being unable to reach him — nor responsibility qua onus — there is no onus upon me to save him, no appropriate ought claim of the general form “I/she/you ought to save the drowning man.” When I was on the dock, however, and I was first witnessing his fall and coming to the conclusion that my boat was fast enough that I could reach him, I had the thought, “I have the responsibility to get into my boat and head out onto the lake to save him.” Was I, at that time, mistaken? Was that thought incorrect? Intuitively, I do not believe I would have been incorrect, but, if this responsibility is the responsibility of accountability, this seems in contradiction

with the premise that I can only be blameworthy for those things that it is genuinely possible for me to do; a premise to which I, as a hard incompatibilist about freedom and accountability, ardently subscribe. Was there, however, still an *onus* on me to do so? Would the claim “I ought to save him rather than continuing to read my novel”, still be *correct* or was I mistaken when it crossed my mind? I believe that there was an onus on me to save the drowning man, and that I would not have been in error to think so. In order to support this, I will say a little about the various forms of responsibility, and this sense of onus, specifically.

### A Variety of Senses of Responsibility

The word “responsibility” can be used in the senses of an assigned task or having been put in charge of something, of a character trait similar to “dependability”, of the feature of having something in your care, of the feature of being the one who will suffer the consequences of some kind or another (as with legal responsibility), and so on. The following are some examples of these senses in use:

- “Your responsibilities will include the sorting and distribution of the morning mail.”
- “I’m not surprised this was not completed on time; he is not very responsible.”
- “She is responsible for the troops under her command.” or “His legal guardian will be responsible for him until he is no longer a minor.”

- “You are responsible for the costs of the damages caused by your dog when he got off his leash.” or “The coach is responsible for the way the team performs.”

Some of these senses include a certain kind of non-moral accountability, whether or not the actor is also morally accountable. To refer to this sense while preventing it from being confused with the moral, causal accountability sense, I will call the former “responsibility qua non-moral answerability” whenever reference to it is necessary. These senses can sometimes appear in combination with morally relevant senses, but they are not inherently morally relevant. There are, on the other hand, senses of responsibility that are inherently moral. These, I argue, are responsibility qua accountability and responsibility qua onus.

Responsibility qua accountability is the kind of responsibility that is tied to praiseworthiness and blameworthiness. There are a number of differing opinions on what is required for this sense of responsibility. I share Derk Pereboom and Galen Strawson’s assessments of the requirements of responsibility of this kind, and agree with their claim that the only way one can reasonably be morally responsible in this sense is to be the ultimate cause of one’s actions. Given that, as a hard incompatibilist about freedom, I do not believe this is possible, I do not believe this kind of responsibility can exist.

The other morally relevant sense of responsibility, however, responsibility qua onus, survives this challenge. When I have a responsibility in the sense of onus, there is an onus on me to perform or refrain from performing certain future actions or to halt or continue actions in progress. When I say that I have a

responsibility qua onus to do this or that, I am saying that I believe that there is an action that is, for all I know, possible, that it satisfies the conditions for right action of the normative theory to which I subscribe, and that it is therefore appropriate to claim that I ought to perform this action. This sense reveals itself in sentences like, “I have a responsibility to do what’s right” or “She is responsible for ensuring her purchases do not promote or condone harmful environmental practices.” I argue that we frequently use “responsibility” in this sense, and that this sense is perfectly coherent with determinism, indeterminism, and the lack of regulative control.

#### “Ought’ Implies ‘Can’”, the Variety of Oughts, and the Unknowable Future

In one of his works, Ted Honderich says, “Taking determinism to be true, as we know, is far from taking the future to be predictable by us” (Honderich 116). Although this is in the context of discussing the enjoyment of life and life goals without free will, I believe it applies here, as well. Even if the future is determined and our actions and their outcomes are fully settled in advance, we experience ourselves as having and making real choices, as deliberating with sincerity, and as having a certain kind of participation in the process. We form judgments about what we ought to do, and we may have a sense of having an onus to act in accordance with those judgments. I experience myself as having a certain kind of responsibility to behave in a given way; not that I will be blameworthy if I do not, but that there is an onus on me to do so, that I am the target of an appropriate ought claim regarding the requirements on me to do so.

Someone might say that some of the objections against moral responsibility qua accountability can be successfully levelled against moral responsibility qua onus. The most seemingly threatening being that if the fact that “ought” implies “can” means that we cannot hold someone accountable for something they could not help but do, because we could not say that they ought to have done it, then it should also be the case that we cannot say that they ought to do something, now or in the future, which they cannot. If “ought” implies “can”, it appears that it cannot, then, be the case that someone ought to  $\phi$  if it is already determined that they cannot, or if it will turn out that, due to indeterministic factors outside of their control, they cannot. And this, you might say, holds true whether we already know this about their future or not. As a result, there can be no onus on them to do any of the things that they cannot, and they could, at most, have this kind of responsibility to do only those things that they are already determined to do, or that it will turn out that they will do. I believe, however, that a closer examination of “ought”, as well as “‘ought’ implies ‘can’” will show that the objections against accountability cannot also be directed at onus.

I argue that not only is there more than one morally relevant sense of responsibility, but also that there is more than one sense of “ought” and, more importantly, there is more than one *morally relevant* sense of ought. There are predictive oughts, such as, “Given that his flight’s departure was delayed by two hours, the landing ought to be significantly delayed.” There are the oughts of hypothetical imperatives: “If you want to prevent yourself from getting caught

shoplifting that sweater, you ought to find a way to remove the security tag discretely.” Additionally, there is the wishful ought, such as in the sentence, “That volcano ought never to have erupted.”<sup>28</sup> These are expressions of a desire for things to have gone differently, or claims about how some good could have been maximized, or other such claims that aren’t relevant to the responsibility of the agent. These three varieties of ought claims are not morally relevant and they are not connected to moral responsibility.

Of the morally relevant oughts, there is the ought found in a proposition like, “He ought to have helped that injured child.” I argue that this kind of thick, past-referencing “ought” fully implies “can”. In other words, I argue that claims with formats similar to “S ought to have  $\phi$ ed” implies “S could have  $\phi$ ed”. As a result, if, due to the lack of regulative control, one could not have done what someone is claiming they ought to have done, that ought claim fails.<sup>29</sup> On the other hand, when someone says “I ought to jump in the boat and save the drowning man on the other side of the lake”, this future-regarding “ought to” implies “can, for all I know”.

In *On What Matters*, Derek Parfit discusses two moral senses of “ought”: decisive-reasons-implying and evidence-relative (Parfit 33-37). The decisive-reasons-implying sense of “ought” is dependent on all of our reasons, whether known or unknown to us (Parfit 33). We may have, for example, reasons to avoid driving home on a given road, because an accident of which we are unaware has

---

<sup>28</sup> This is usually phrased in a way that resembles “That volcano should never have erupted”, but I would still consider this to be an ought claim, for reasons there is not room to discuss here.

<sup>29</sup> As seen above, there are times when propositions with the “S ought to have  $\phi$ ed” format do not fail due to being of a wishful variety, rather than of a normative kind. Ought claims of the wishful variety are not related to responsibility.

just blocked a bridge along that route. It seems perfectly sensible to say that you have these reasons, though they are unknown to you. These are our all-things-considered reasons, from a god's-eye view.

Parfit's evidence-relative sense of "ought" is based on the reasons that we have, given the facts that we have. It is what we have the most reasons to do, considering the evidence at our disposal. It is on these reasons, Parfit tells us, that we must base our decisions, seeing as we cannot know all of the facts applicable to our decisions (Parfit 36-37). I argue that Parfit's conception of the evidence-relative ought is similar in some respects to the kind of ought that I propose retains its strength regardless of free will and regulative control. His is a kind of "ought" similar to that of which we are making use when we say that we ought to do this or that, in a way we intend to carry normative weight.<sup>30</sup> If I have evidence that I have no regulative control, I might have reason to believe that I am no longer bound by any ought whatsoever and this evidence of my lack of regulative control might become a piece of evidence shaping my beliefs that there is nothing in particular that I ought to do, giving me reasons to do nothing in particular. I think this may be true of an evidence-relative sense of an accountability-implicating "ought", but not of an onus-implicating "ought". Evidence-relative senses of "ought" imply only evidence-relative senses of "can". Once an action has taken place we have all of the evidence relative to what one ought to have done, because we have all of the evidence relative to what they could have

---

<sup>30</sup> It is not identical to the kind of ought I propose, however.

done.<sup>31</sup> Before the fact, however, we do not have all of the evidence about what one can do. We do not have evidence that there is nothing that we ought to do or nothing that we can do, we simply do not have evidence about what we *will* do. We have our evidence about the facts relative to the decision in question, and we have limited evidence about which of the options may conceivably be open to us, but the facts about what we *can* do are obscured from that perspective. Seeing as we are not omniscient, only ought claims that take into account the facts that appear available to us are appropriate.

“Ought to” Implies “Can, for All I Know”

I believe that such future-regarding ought statements contain a kind of unarticulated assumption about what can be done, even if the individual making the ought claim has not reflected on these assumptions. Elsewhere, I have put it as follows:

“A statement containing an ‘ought’ in this evidence-relative sense contains unarticulated constituents that are contingent upon the evidence held by the utterer (or the believer, if the belief is not uttered). In other words, assuming this sense, ‘S ought to X’ expresses, ‘Given the evidence that I have, S ought to X.’ and it implies, ‘Given the evidence that I have, S can X.’”<sup>32</sup> (Renaud 4)

If we return to the example of my dilemma on the pier, I believe that the content of the ought claim, “I ought to get in my boat and save the drowning man” can be expanded to, “I ought to get in my boat and save the drowning man because, for all I know, I can.” I am not arguing that this is the represented

---

<sup>31</sup> I recognize that this assumes the correctness of hard incompatibilism about freedom. I am working under this assumption until I have evidence-relative reasons to do otherwise.

<sup>32</sup> I would now refrain from calling them unarticulated constituents.

mental content of the thought or of the belief; I am arguing that these ought claims entail this further belief that I am capable, for all I know, of completing the given activity. These forward-looking “ought” propositions can be true even if it is the case that we will not, and even cannot do the action in question due to our lack of relative control. It is for this reason that I think that we can still have a responsibility in the sense of having an onus, even if determinism is true. There can still be an action that is, for all I know, possible, that satisfies the conditions for right action of the normative theory to which I subscribe, and there can still be an onus on me to perform this action. That is, there is an action that I ought to perform; and this, regardless of whether it will turn out that I actually can.

### The Role of Evidence

Ought claims of the “I/You/S ought to  $\phi$ ” format can be made from the first-, second-, or third-person perspectives and I believe they can be further expanded as, “Given the evidence that I have, which includes the evidence that my/your/S’s  $\phi$ ing is possible, “I/You/S am/are/is morally required to  $\phi$ .”

It strikes me as though these claims, when made from the second- or third-person perspectives, could only reasonably initiate an onus, a moral requirement, on S if they included a conditional clause about the evidence of which S is in possession. Although it is quite common for us to declare that someone ought to  $\phi$ , even if they do not know they ought to  $\phi$ , I argue that we do not tend to mean that there is a genuine expectation or requirement placed upon S to  $\phi$  if S does not have the evidence that makes it appear as though S ought to  $\phi$ .

Let us assume I am watching a news broadcast and become informed of a breaking recall of a medication currently prescribed to my neighbour's young daughter. I know my neighbour has not seen the broadcast, as I can see through the window that he is in his garden. I may think to myself, "He ought to call his clinic and have his daughter's prescription changed." I believe this is the best thing for him to do, and so I believe that he *ought* to do this, but I believe he ought to do it in the same way as I believe the charitable organization ought to win the pick-7 lottery. That is, I believe that there is some good that could be protected or maximized by the event taking place, and so this is a wishful "ought". I do not believe that there is an *onus* on my neighbour to stand up straight away, march into the house, and call his doctor, given that I know he is not in possession of the information I have. It seems too unreasonable to require that someone act in line with evidence they do not (yet) have.

This onus can only be present if the moral requirement is reasonable, and the moral requirement is reasonable only if it is based on the evidence possessed by the individual expected to perform or refrain from performing the action in question. In other words, I think that, "She ought to purchase items with reduced packaging as it is better for the environment." can be expanded to "Given the evidence that I have, she is morally required to purchase items with reduced packaging as it is better for the environment, if she has or comes to have the same evidence as I now have." If I generalize these expansions, they follow a format similar to the following:

First-person: “Given the evidence that I have, I am morally required to  $\phi$ .”

Second-person: “Given the evidence that I have, if you have (or acquire) the evidence I have, you are (or will be) morally required to  $\phi$ .”

Third-person: “Given the evidence that I have, if S has (or acquires) the evidence I have, S is (or will be) morally required to  $\phi$ .”

Although the content of a second- or third-person “ought” claim depends on the evidence and beliefs held by the utterer or believer, not by S, a reasonable moral requirement can only be generated and apply to S if S, too, has the relevant evidence. In other words, a moral requirement in connection to an ought claim exists only if the utterer and the actor share generally the same beliefs about the evidence.

#### Appropriate Onus-implying Ought Claims

For an onus-implying ought statement to be appropriate, it must be logically possible according to the evidence possessed by the one making the claim. In other words, it must respect “‘ought’ implies ‘can, for all I know’”. An appropriate onus-implying ought must also be reasonable. Admittedly, the idea of reasonableness is vague and somewhat loaded. Here, I intend it to mean only that one should not be expected to act on evidence-relative reasons if one does not possess the same reason-generating evidence. I should not expect evidence that I have and you do not to generate evidence-relative reasons for you (unless

and until I convey that evidence to you or you have acquired it). In this way, it hints at something closer, perhaps, to “reason-able” than “reasonable”, in that the acting individual is able to generate the same evidence-relative reasons as the person with the ought claim in mind.<sup>33</sup>

As it is not reasonable to expect someone to act on evidence-relative reasons for evidence they do not have, these ought claims are only reasonable if and when they can be expanded into a proposition that includes a conditional clause about the would-be-actor’s beliefs. They are, therefore, reasonable, only if they are in a format similar to those mentioned above; in other words “Given the evidence that I have, if you/S have/has (or acquire/acquires) the evidence I have, you/S are/is (or will be) morally required to  $\phi$ .” An ought claim that did not expand into one of this general form, one that claims that an individual has a moral requirement to do something for which she may have no evidence or evidence-relevant reasons is not reasonable and does not make appropriate use of an onus-implying “ought” claim. First-persons-relative ought claims do not require this conditional clause, as the actor then has the same evidence and evidence-relative reasons as the person making the claim.<sup>34</sup>

### Onus Remains Unharmful by the Lack of Free Will

Ought statements that have the above features, those that are logically possible and reasonable, are appropriate and the target of these appropriate

---

<sup>33</sup> Perhaps it is only necessary that the acting individual generate nearly the same evidence-relative reasons, or evidence-relative reasons that support the same action, rather than perfectly identical reasons.

<sup>34</sup> At least, this is true in the most typical cases. Perhaps we could argue that cases that include dissociative episodes, multiple personality disorders, amnesias, and other such cases will require some kind of conditional, as though working from a second- or third-person perspective.

onus-implying ought claims are morally responsible, in the responsibility qua onus sense. To be the target of an appropriate onus-implying ought claim is to be morally responsible, in this sense. This sense of moral responsibility, as you can see, would be unharmed by the existence of determinism, indeterminism, or any of their entailments for free will. Therefore, even if determinism is true, incompatibilists and hard incompatibilists are correct, and there is no regulative control, we need not abandon the normative practice of claiming that one ought to do this or that, as this practice is grounded in a different sense of “ought”, one that remains coherent without incompatibilist freedom. We are still, in at least this sense, morally responsible for continuing to act according to our guiding, normative principles.

#### Responsibility Hemi-incompatibilism

I call this view that there exists a forward-looking responsibility qua onus regardless of the existence of free will, but no responsibility qua accountability, “responsibility hemi-incompatibilism”. The “hemi” is similar to that of “hemisphere”, which points us to one half of the globe; to one side of the equator or another. By using this term, I intend to indicate that moral responsibility is likewise divided by an epistemic equator. On the one side, we have events and states that have not yet come about, and to which we have no knowledge, the unilluminated side. On the other, sits the past. I argue that we can attribute a kind of responsibility, “onus”, to moral agents for actions that are on this unilluminated side of the epistemic equator. Note that although the two sides of the epistemic equator are currently coextensive with the two sides of a moving chronological

equator drawn always at the present time, I concede that I cannot be certain that this will always be the case. If universal indeterminism is true, it seems likely that it will have to be and if determinism is true, there is, I suppose, the possibility that we will gain epistemic access to future events and states. The epistemic equator can break apart from the chronological equator without problems for this account. As a result, time travel and fortune teller counterexamples will not harm the basic argument that moral responsibility, qua onus, can only be attributed to actions on one side of the epistemic equator.

Responsibility hemi-incompatibilism is an included-agent account of moral responsibility. The criteria for responsibility include features of the agent and what the agent knows, as well as things about the external world and what is possible. It is also an autonomous account, as it does not rely on normative claims or the endorsement of any normative accounts. This view captures, I believe, a sense of moral responsibility that is often ignored in the free will debate, and separates it from the sense that is threatened by the lack of regulative control.

## 5. Similar Views

### Bruce Waller, *Against Moral Responsibility*

In his book *Against Moral Responsibility*, Bruce Waller expresses a hard incompatibilist stance about moral responsibility, which he uses in the sense that I would call responsibility qua accountability. He, too, thinks that we cannot have a deserts-based moral responsibility, believes that “the moral responsibility belief system is fighting a running retreat against scientific research that renders the system less and less plausible”, and claims that it should be abolished “root and branch” (Waller 1-2). He also, like me, believes that our moral judgments and our use of ought claims, at least in certain circumstances, are consistent with the denial of regulative control and desert-based responsibility and that we can continue to discuss the rightness and wrongness of actions without responsibility, regulative control, and blame (Waller 3, 179-181, 185). I find in Waller what appears to be an ally, and our conclusions seem to me more alike than they are different. There are still, however, a number of points on which we differ.

### Waller’s Senses of “Ought”

Waller and I both maintain that there are different moral senses of “ought” and that their disambiguation allows us to see how we can retain certain types of moral judgments without desert-based responsibility, responsibility qua accountability.<sup>35</sup> Waller claims there are *admonition oughts* and *judgment oughts* (Waller 187). An *admonition ought* is used to reprove an individual for their

---

<sup>35</sup> Although he does not mention predictive oughts or the oughts of hypothetical imperatives, there is no reason to think he would exclude them as legitimate senses of “ought”; they were merely non-relevant to the discussion about moral oughts.

behaviour. It is a type of ought that involves a certain amount of blame and is, he believes, pointless if the object of the admonishment has no regulative control or desert-based responsibility (Waller 187). The *judgment ought*, however, expresses that one has noticed a “moral flaw”, without the expectation that the individual could have changed the behaviour being judged. This variety of ought involves a kind of commiseration about something that cannot be changed, as we might commiserate over the diagnosis of a serious disease, and provides moral education to those present (Waller 188). Waller believes that denying moral responsibility does not harm the *judgment ought*, which he thinks serves a useful and important purpose. He claims that, in order to see the legitimacy of this second kind of ought, one must reject the principle of “ought implies can” (Waller 189).

Although our foundations are similar, and we both assert that oughts are not harmed by the denial of regulative control and responsibility qua accountability, our lines are drawn in different locations. Waller lumps what I call onus-implicating oughts and accountability-implicating oughts together under the category of *admonition oughts*, which may be a mistake. While I think his *admonition oughts* capture the rebuking or blaming function of certain types of oughts, I think those oughts are the “ought to have done”, accountability-implicating oughts. To put these in the same category with the “ought to do”, onus-implicating oughts fails to capture, highlight, and isolate the judgment that we are making that the target of the ought has an obligation to do (or refrain from doing) an action in question.

One might want to say that the *judgment ought* covers this obligation component, but *judgment oughts*, as they are described, are not strong enough to do so. They merely express a judgment that something would be good or preferable to something else. *Judgment oughts* are wishful oughts and do not seem to include the strong imperative, the moral obligation to do something or other, that my onus-implying ought does. I think this is lost, prematurely, in this division.

If *judgment oughts* are meant to encompass commiseration about things like tornados and diseases and whether they ought to be, I believe the category of wishful oughts does a better job of the task of making clear just what is expected, and what should be expected, from this kind of ought. The wishful ought is not laced with the moral judgment of (some) *judgment oughts*; this is not a failing of the category, but rather the redrawing of lines precisely to disambiguate those times where a moral judgment is being made from the times where one is not.

By separating out the moral judgment from the *judgment ought*, and likewise dividing the *admonition oughts* into onus-implying and accountability-implying oughts, we can see that the features of the oughts discussed by Waller are better served by three categories of ought than by two. The wishful ought captures desires, commiseration, preferences, judgments about what is good, etc. The accountability-implying ought captures the blaming portion of the *admonition oughts* that Waller and I believe are nonsensical without regulative control. Finally, the onus-implying ought captures the judgment about what we believe

would be morally correct for someone to do if they are in the position of the target of our onus-implying ought.

Additionally, the onus-implying ought captures something that I believe Waller's two categories of oughts fail to capture. When we say to someone that they ought to do something, we are saying not only that we think it would be preferable that they do, we are saying that we believe that, if it is possible for them to do so, there is an *obligation* for them to do so. When I tell someone that they ought to refrain from stealing the car parked on the street from the driver who left their keys in the ignition, I am not merely indicating that I believe this is the optimal situation for which we should hope. I am indicating that I believe they are obliged, should they be capable of doing so, to refrain from it. I can refer only to the way I play the language game along with the way I perceive others as using these terms as evidence for my claim, though empirical studies about what one means when one uses the terms could help illuminate this if this is indeed how "ought" is used and what it means in cases such as these. And, although I am convinced that terms should be corrected and given new, revisionist definitions or, at times, abolished altogether if their use is incoherent, I do not believe this is the case for onus-implying oughts. As I have claimed above, there is ample room for their continued use in a (prospective) paradigm in which we have rejected regulative control.

#### Waller and the Rejection of "Ought" Implies 'Can'"

Waller believes we do not have good reasons to think that "ought" implies "can" (Waller 185). He thinks that our support for this claim comes from the

assumption that should we fail to do something that we ought, or should we do something that we ought not, we would deserve blame. If we do not deserve blame for things we cannot do, and if we deserve blame for things we ought to do that we fail to do, then it cannot be that we ought to do things that we cannot.

The incorrect argument would look something like the following:

- 1) If we cannot  $\phi$ , then we do not deserve blame for failing to  $\phi$
- 2) If we ought to  $\phi$ , then we deserve blame for failing to  $\phi$
- C) If we cannot  $\phi$ , then it is not true that we ought to  $\phi$

Waller takes exception with the second premise. He believes that if we reject the assumption that we deserve blame for failing to  $\phi$  if we ought to  $\phi$ , then we can also reject “‘ought’ implies ‘can’” (Waller 183). He thinks we need to recognize that the world is not constructed such that “every good thing can be accomplished; the world is not set up to satisfy our moral demands or desires” (Waller 183).

Waller appears, however, to show some uncertainty about just how completely we should reject “‘ought’ implies ‘can’” and perhaps to accept it in some form or on some level. He tells us that,

“After all, it may make perfectly good sense to say that I ought to rescue a child stuck in a tree, because I have the capacity to reach up and rescue the child; however, it would be nonsense to suggest that I ought to rescue people falling toward Earth in a 747. With the best will in the world, I can’t accomplish it. Perhaps Superman ought to, but not me.” (Waller 185)

He also says, while comparing the inability to stop a 747 from crashing with the inability to stop being jealous,

“[...] as we watch the doomed 747 plunge toward destruction, you say to me, “You ought not let that plane crash.” Here the “ought” statement really is false; indeed, I shall have trouble even making sense of your statement. I share with Sam an incapacity to perform the action in question: he cannot stop being jealous, and I cannot rescue the airliner. But that common capacity does not lead to a common result: the “ought” statement addressed to me is false, or perhaps nonsensical; the “ought” statement addressed to jealous Sam is true, useful, and quite intelligible. Thus, when it is not true that one can, it does not automatically follow that one is not a proper object of “ought” language. The difference is that no amount of moral resolve or proper conditioning or ethical admonition will make me into someone capable of rescuing malfunctioning jetliners. But when you say that I ought to avoid jealousy, or I ought to work harder at teaching, or I ought to stop smoking, that is to admonish me to do things that are within the capacities of at least some humans.” (Waller 186)

Without the claim that “‘ought’ implies ‘can’”, however, it is hard to explain why it is nonsensical to say that S ought to stop the 747 from crashing. S’s inability to do so clearly influences the requirement for S to do so. The difference between the statement that S ought to prevent the crash and that S ought to stop being jealous or that S ought to quit smoking is that we know, with as near complete certainty as we can have, that S cannot stop that airliner’s free-fall. We do not have the same certainty that S cannot, in fact, stop being jealous or quit smoking. S can, for all we know, do either of those things. If we were truly certain that S could not bring an end to his jealousy, perhaps because we had implanted a chip in S’s brain that would continuously stimulate neural connections that

foster jealousy, I think we would no longer claim that he *ought* to do so; at least not in the same way as we would otherwise. We might still say it in the wishful sense, but this is an ought of a very different colour.

Saying, instead, that “ought to have done’ implies ‘could have” but that “ought to do’ implies ‘can, for all I know” captures the reason we can say that S ought to stop smoking, but that we cannot say that he ought to save the 747 airliner.

Waller also claims that if I borrow money from you but cannot repay you because of a personal financial crisis, it is ridiculous to say that I no longer ought to pay you. He offers the fact that it is still true that I ought to pay you, despite the fact that I cannot as further evidence that “ought” does not imply “can” (Waller 184). I believe that this is merely a case of having under-specified the details of the ought claim in question. It is not that I no longer ought to pay you back, it is that it is not sensible to say that I ought to pay you back this very minute, when I cannot. It is still true that I ought to pay you back in the future, when I again can<sup>36</sup>.

### Usefulness

Waller says that it may be useful, for the moral education of those present, to say that Sam ought to stop being jealous, even if we know that Sam, himself, cannot (Waller 186). But here again, I believe this is a wishful ought. We are indicating that it would be preferable if it were the case that Sam be able to temper his jealous tendencies. This expression of a judgment that the alternative scenario would be preferable provides, I believe, an equally effective role in the

---

<sup>36</sup> Perhaps it is also even true that I made mistakes that unnecessarily or neglectfully led to the crisis that prevented me from paying you, but this is irrelevant to the onus-implying ought claim.

moral education of bystanders. That claim is empirically verifiable and, even should it turn out to be incorrect, the truth of whether Sam ought to do something does not rest on the usefulness of saying so.

He also says that it may be useful to say that Sam ought to stop being jealous because it may nudge Sam, himself, towards taking the necessary steps to achieve this goal (Waller 186). Like the case of the moral education of bystanders, its usefulness does not make it true. In this case, however, there is additional forcefulness behind the ought claim, if the ought being conceived is an onus-implying ought, as we are not merely saying that it would be preferable for Sam to not be jealous, we are also saying that if he can take steps to achieve that goal, there is an onus on him to do so.

#### Against Some Kinds of Responsibility

Despite these differences, there is a great deal of agreement between Waller's views and my own. The key disagreement may be that he calls moral responsibility only those things that are connected to what I call responsibility qua accountability, whereas I believe that the onus component of moral responsibility that is not present in Waller's view is a real and important sense of moral responsibility; one that remains unthreatened, despite the denial of the other factors he and I both reject.

#### Derk Pereboom, "Free Will Skepticism, Blame, and Obligation"

In his draft of the forthcoming chapter "Free Will Skepticism, Blame, and Obligation", Pereboom describes the type of responsibility that I call responsibility qua accountability and says that this is the kind about which we typically talk

within the context of the free will debate. He tells us, however, that both “moral responsibility” and “ought” are used in several different ways (Pereboom 1, 6). He believes, as do I, that some of these other senses of moral responsibility are not threatened by the lack of regulative control and can be affirmed by those who deny free will (Pereboom 1).

### Oughts of Axiological Evaluation and Oughts of Specific Agent Demand

Although the main task of his chapter is to defend a revisionist type of blame, part of the discussion is devoted to the examination of different kinds of ought claims, as well as to the “‘ought’ implies ‘can’” principle. Pereboom tells us that there are some senses of “ought” that are not linked to this principle (Pereboom 5). Like me, he distinguishes between the way we think things ought to be, which I call wishful oughts, and oughts that are about what people ought to have done, or ought to do. He calls the former type *oughts of axiological evaluation* or of *axiological ideality* (Pereboom 6). He believes that these types of ought claims do not imply “can”. He maintains, however, that “ought to do” claims, which he calls *oughts of specific agent demand*, do invoke the “‘ought’ implies ‘can’” requirement. He explains that the reasonableness of claims that use this sense of ought is threatened by determinism, that it would be unfair to claim that an agent who could not have performed a given action ought to have performed that action (Pereboom 6). He adds that, on the other hand,

“it might well not be unfair for me to recommend to that agent that she perform an action of that type at some future time, given that it is epistemically open that she will develop the requisite motivation by then, and in particular if the recommendation might plausibly contribute causally to producing the

motivation. And to recommend the action to her, I might tell her that she ought to perform the action at the future time, and do so appropriately and without making any kind of mistake” (Pereboom 6).

### Oughts of Axiological Recommendation

Pereboom thinks that the particular kind of ought we employ when we are recommending future action, what he calls an *ought of axiological recommendation*, is different from the ought of specific agent demand, and that it is a subvariety of an ought of axiological evaluation (Pereboom 6). Because this type of ought, like other oughts of axiological evaluation, does not require that there be “a route accessible to an agent, via reasons for action, to her acting in some relevant way”, this type of ought, he thinks, is still perfectly compatible with determinism (Pereboom 6-7).

This particular kind of ought looks, up to this point, to be much like my onus-implicating ought. It differs, however, in two key ways. The first of these is that it is, above all, an evaluation. Oughts of axiological recommendation are judgments about agents’ actions, but these oughts do not contain the implication of an obligation to act or refrain from acting (Pereboom 16). They, like Waller’s *judgment oughts*, highlight what would be optimal, but they fail to capture the onus-implicating component of our claims about what one ought to do. I maintain that whether or not there is “a route accessible to the agent” for performing an action, that the possibility of the agent performing that action remain “epistemically open” is, for the reasons I have already discussed above, enough to make this type of ought coherent and reasonable.

The second way in which oughts of axiological recommendation differ from my onus-implicating oughts is that, although for Pereboom oughts of axiological recommendation are meant to make recommendations about future behaviour to an agent, they can sometimes have past actions as their targets. In other words, I can say that S ought not to have  $\phi$ ed as a way of suggesting that S no longer perform actions of the type  $\phi$  (Pereboom 7-8). This breaks significantly, I believe, with common use of ought claims about past actions. As I have already indicated, I have little desire to resist the reassignment of words to new meanings if there is adequate reason to do so, but I fail to be convinced that the change is warranted here. To say, “you ought not to have done that” when what one means is that “you ought not to do that again in the future”, and the latter is an available option that preserves senses of “ought” already in use, is unnecessary and confusing. It seems less unclear to simply retire the use of the accountability-implicating sense of ought entirely if we have no regulative control, preserving the surviving onus-implicating sense and using its forward-looking standpoint.

For these reasons, although I am in agreement with much at the core of both Waller’s and Pereboom’s accounts, I believe my alternative account for the division of ought claims and the sense of moral responsibility that can survive determinism and other threats to regulative control remains a better alternative.

## 6. Conclusion

Over the course of this project, I have attempted to show that Helen Steward's new libertarian alternative, even if it could be shown to be a metaphysically viable explanation of a kind of agent causation, cannot satisfyingly ground moral responsibility qua accountability and does not, therefore, do enough to call into question the hard incompatibilist rejection of desert-based responsibility.

I have also attempted to present an outline of my own responsibility hemi-incompatibilism. This account, I believe, captures our current uses of the terms related to moral responsibility, and discards only those that are truly threatened by the lack of regulative control. Additionally, it is able to explain how there is room for normative prescriptions and proscriptions in a world without regulative control, and that we need not throw out the ethical baby with the bathwater.

There are those who would claim that it is impossible to preserve any kind of moral orientation having rejected the senses of "responsibility" and "ought" that I propose we abandon. I maintain, however, that it is not only possible, but also highly desirable. Along with Waller and Pereboom<sup>37</sup>, I believe that it is not only possible for us to continue benefiting from the goods provided by interpersonal relationships and a rich moral life, but that both of the latter would be better served by clearly seeing the limits of our agency, and by beginning to adjust our social and interpersonal interactions and institutions accordingly.

---

<sup>37</sup> See Waller 190-199 and Pereboom 8-11.

## Works Cited

- Dennett, Daniel C. *Elbow Room: The Varieties of Free Will Worth Wanting*. Oxford: Oxford University Press, 1984. Print.
- . *Freedom Evolves*. London: Penguin Books, 2003. Print
- Ekstrom, Laura Waddell. *Free Will: A Philosophical Study*. Boulder, CO: Westview Press, 2000. Print.
- Fischer, John Martin. *My Way: Essays on Moral Responsibility*. Oxford: Oxford University Press, 2006. Print.
- . "Précis of My Way". *Philosophy and Phenomenological Research* 80:1 (2010): 229-241. Wiley Online. Web. 19 January 2012.
- Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person." *The Journal of Philosophy* 68:1 (1971):5-20. Jstor. Web. July 18, 2013.
- Haji, Ishtiyaque. *Incompatibilism's Allure: Principal Arguments for Incompatibilism*. Peterborough, Ontario: Broadview Press, 2009. Print.
- Honderich, Ted. *The Determinism Problem*. New York: Oxford University press. 1993. Print.
- James, William. "The Dilemma of Determinism." *The Works of William James: The Will to Believe and Other Essays in Popular Philosophy*. Cambridge, MA: Harvard University Press, 1979: 114-140. Print.
- Kane, Robert. *A Contemporary Introduction to Free Will*. Oxford: Oxford University Press, 2005. Print.
- O'Connor, Timothy and Hong Yu Wong, "Emergent Properties", *The Stanford Encyclopedia of Philosophy*. Spring 2012 Edition. Ed. Edward N. Zalta.

- Stanford University, 2012. Web. 6 June, 2013.
- <<http://plato.stanford.edu/archives/spr2012/entries/properties-emergent/>>.
- Parfit, Derek. *On What Matters*. Vol. 1. New York: Oxford University Press, 2011. Print.
- Pereboom, Derk. "Free Will Skepticism, Blame, and Obligation." *Blame: It's Nature and Norms*. E. Neal Tognazzini and Justin Coates. New York: Oxford University Press, forthcoming. Electronic draft version.
- . *Living Without Free Will: The Case for Hard Incompatibilism*. The Oxford Handbook of Free Will. Ed. Robert Kane. New York: Oxford University Press, 2002, 477-488. Print.
- Renaud, Mary. *An Account of Moral Responsibility that Is Compatible with Determinism*. Coursework presented to Professor David Matheson. Carleton University, 2012. Print.
- Steward, Helen. *A Metaphysics for Freedom*. Oxford: Oxford University Press, 2012. Print.
- Strawson, Galen. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75:1/2 (1994): 5-24. Jstor. Web. 17 June, 2013.
- Strawson, Peter F. "Freedom and Resentment." *Freedom and Resentment and Other Essays*. London: Routledge, 2008. 1-28. Print.
- Van Inwagen, Peter. *An Essay on Free Will*. Oxford: Clarendon Press, 1983. Print.
- Waller, Bruce. *Against Moral Responsibility*. Cambridge, MA: MIT Press, 2011. Print.