

2002

# Free will and responsibility: indeterminism and its problems

Troy Dwayne Fassbender

*Louisiana State University and Agricultural and Mechanical College, tfassb1@lsu.edu*

Follow this and additional works at: [https://digitalcommons.lsu.edu/gradschool\\_theses](https://digitalcommons.lsu.edu/gradschool_theses)



Part of the [Arts and Humanities Commons](#)

---

## Recommended Citation

Fassbender, Troy Dwayne, "Free will and responsibility: indeterminism and its problems" (2002). *LSU Master's Theses*. 2705.  
[https://digitalcommons.lsu.edu/gradschool\\_theses/2705](https://digitalcommons.lsu.edu/gradschool_theses/2705)

This Thesis is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Master's Theses by an authorized graduate school editor of LSU Digital Commons. For more information, please contact [gradetd@lsu.edu](mailto:gradetd@lsu.edu).

FREE WILL AND RESPONSIBILITY: INDETERMINISM AND ITS PROBLEMS

A Thesis

Submitted to the Graduate Faculty of the  
Louisiana State University and  
Agricultural and Mechanical College  
in partial fulfillment of the  
requirements for the degree of  
Master of Arts

in

The Department of Philosophy and Religious Studies

by

Troy Dwayne Fassbender  
B.A., Louisiana State University, 1999  
May 2002

# Acknowledgements

I would like to thank the faculty of the Department of Philosophy and Religious Studies at Louisiana State University, especially the members of my thesis committee. I could not recommend more highly to other graduate students than that they enroll in any course taught by Husain Sarkar and that they choose him as their thesis director. His sharp mind and generous spirit have been a boon both to me and to my work. I would like to thank the rest of my committee, Jon Cogburn and John Baker, for their invaluable insight and recommendations. Parts of the second chapter were presented to the Alabama Philosophical Society Conference in 2001, and my arguments benefited from the question and answer session that followed my presentation. Finally, I thank my wife Joyce for her proofreading as well as her good cheer on those bleak nights when I would blankly stare at my computer with a knotted belly and a fevered mind.

# Table of Contents

Acknowledgements .....	ii
Abstract .....	iv
1. Introduction .....	1
1.1 Incompatibilists .....	2
1.2 Compatibilists .....	4
1.3 Looking Ahead .....	5
1.4 End Notes .....	8
2. Robert Kane's Incompatibilism .....	9
2.1 Compatibility and Significance .....	9
2.2 Intelligibility and Existence .....	25
2.3 Criticisms .....	32
2.4 Concluding Remarks .....	44
2.5 End Notes .....	44
3. Four Problems for Indeterministic Accounts of Freedom .....	47
3.1 The Strawson Challenge – No Place for Indeterminacy .....	47
3.2 Nagel's Problem of Autonomy .....	56
3.3 Dennett's Compatibilist Shift .....	61
3.4 Objection from Rational Explanation .....	68
3.5 Concluding Remarks .....	73
3.6 End Notes .....	73
4. The Limits of Indeterministic Freedom .....	75
4.1 Introduction .....	75
4.2 Incompatibilism and Beta-Like Rules .....	77
4.3 Implications of Beta .....	84
4.4 Beta and Moral Responsibility – The Classical Tradition .....	100
4.5 Concluding Remarks .....	107
4.6 End Notes .....	108
5. Conclusion .....	110
5.1 Further Problems for Indeterminists .....	111
5.2 Concluding Remarks .....	125
5.3 End Notes .....	126
References .....	127
Vita .....	129

# Abstract

This work is devoted to criticisms of libertarian philosophers who attempt to provide an account of agent freedom that relies solely upon indeterminism. First, the philosophy of Robert Kane is examined. I argue that Kane's account does not succeed as an intelligible libertarian account of freedom and at best makes compatibilist accounts more intuitive. I next examine objections to indeterminist accounts as lodged by Galen Strawson, Thomas Nagel, Daniel Dennett, and Richard Double before turning to an analysis of a debate among Peter van Inwagen, John Martin Fischer and Mark Ravizza. Van Inwagen argues that we are seldom able to do other than we do but as long as we are in some way responsible for this inability then this does not entail that we can only rarely be held responsible. Typical cases are those in which an agent's character determines a particular action and the agent is responsible for having the character she has. Fischer and Ravizza argue that van Inwagen's account is empty because the character of an agent is formed at an early age by forces beyond her control.

I conclude by arguing, *pace* Kane and van Inwagen, that even if an action is determined by an agent's character and the agent is responsible for having that character, we still may not be able to hold the agent responsible in a significant amount of cases. Additionally, I attempt to provide a compatibilist solution to the problem of free will in an attempt to show that the ability to do otherwise is not relevant to the problem of free will.

# 1. Introduction

There is a classical tradition in philosophy that is characterized by a debate between compatibilism and incompatibilism.<sup>1</sup> The former position holds that free will is compatible with the thesis of determinism and the latter argues that it is not. The importance of this debate can be seen by considering the relationship between free will and moral responsibility. In order to hold an agent responsible for a state of affairs that results from an action he performed, it is generally required that the agent must have had a choice concerning whether or not he would perform the action. It must be true that the agent could have done other than what he in fact did. If the resulting action was not one the agent had a choice about performing, we would not tend to hold him responsible for that action. While driving, the brakes in Susan's new car malfunction resulting in her Ford plowing through an intersection and hitting a Datsun. An insurance investigator later determines that the brakes in Susan's car were defective and installed incorrectly by the manufacturer. Because of this Susan should not be held responsible for the action, she had no choice concerning whether or not the car she was driving would hit another.

Most incompatibilists play on this relationship between free will and responsibility to argue that the thesis of determinism must be false. In this vein, incompatibilists have provided several reductio arguments against the truth of determinism. These typically begin by assuming the truth of determinism and then showing it has as a consequence that no one is ever able to do other than what they have done. If this were true, then it would seem that no one could be held responsible for his or her actions. Suppose Susan had known in advance about the faulty brakes yet chose to

drive the Ford anyway. If determinism were true, so the arguments go, we would still not be able to hold Susan responsible for the resulting accident. This is because every choice, including her decision to drive with faulty brakes, is mere illusion. Her deciding to drive the car is akin to a rock deciding that it will fall to the Earth when dropped. Since this conclusion is phenomenologically absurd, determinism is shown (in the minds of incompatibilists) to be false.

In this introductory chapter, I discuss the various positions held by compatibilists and incompatibilists. This discussion will be brief because the positions are dealt with extensively throughout the course of the next several chapters.

### **1.1 Incompatibilists**

Incompatibilist philosophers can be divided into two camps. In the first camp are the hard determinists. The hard determinist agrees that free will is not compatible with determinism but, unlike the other incompatibilists, concludes that free will does not exist. Examples of hard determinist include Thomas Hobbes, Arthur Schopenhauer, J. S. Mill, and Ted Honderich.<sup>2</sup> Hobbes, Schopenhauer and Mill are psychological determinists and argue that an agent is always determined to act by his strongest desire that is in turn determined by his heredity and upbringing. In the second camp are the libertarians. The libertarians are the incompatibilists of the ilk mentioned in the previous paragraph, those that argue that determinism is false. The libertarians can further be divided into two groups. The first group relies upon special types of entities or special forms of causation to gain libertarian freedom. Immanuel Kant and Roderick Chisholm represent this group. The second show a naturalist streak by relying solely upon indeterminism to gain

libertarian free will. Examples of libertarian indeterminists are Robert Kane and Peter van Inwagen.

Libertarians of any stripe argue that there are conditions that are necessary for the existence of free will. One such condition is, generally, that free will cannot exist if agents can never do other than what they in fact do. This is the familiar worry discussed earlier that all choice is but a mere illusion. This worry can be couched in terms of alternate possibilities. If determinism is true, then there is but a single open possibility ahead of us. No alternative possibilities could exist. Because there are no alternative possibilities open to us, it would be true that in every situation we lack the ability to other than what we do.

Generally, some form of allowance is given for actions that are determined by an agent's character. Because of the good character that she has developed, Mother Teresa was unable to turn away from someone in need. Although it is true that in each particular instance of not turning away she could not have done other than what she did do, we can hold her responsible for her actions because she was responsible for forming her character. If an agent is responsible for forming her character, then she is also held responsible for actions that are determined by that character.

Libertarians of the indeterminist variety also tend to shun the libertarians who rely upon special entities or special forms of causation to gain freedom. The indeterminist typically finds these to be unintelligible and argue that their implausibility hurts the libertarian cause more than helps.

## **1.2 Compatibilists**

Though compatibilists argue that free will is compatible with determinism, most do not argue further that determinism is necessary for free will. Rather, free will is compatible with both determinism and indeterminism. However, they would not agree with the incompatibilist that indeterminism can make a meaningful contribution to solving the problem of free will. They argue that no intelligible account of libertarian free agency has been proposed (nor can be). Just because an action is not determined does not seem to make it an action for which an agent can be held responsible. A typical compatibilist criticism of indeterminists is that there is no place that indeterminism can be introduced that would result in actions for which agents can be held responsible. Indeterminism seems to be just chance, and it is hard to see how simple chance aids the libertarian. The picture painted by these critics is bleak. We return to when Susan was deliberating about whether or not to drive her Ford knowing that the brakes do not work. She decides to do the right thing and leave the car in the driveway. However, before she actually does so, indeterminism interferes and alters her decision so that she ends up having the accident. In examples like this the introduction of indeterminism serves as a barrier to freedom and not as an aid to it.

Compatibilists typically equate free will with the freedom of an agent to do whatever it is he wants to do. So long as I am not coerced by others or by circumstance from doing what I desire, I am free. Thus we can distinguish between my walking across a street because it is what I desire and my being forcibly blown across the same street by a strong wind. In the first case, I was free. In the latter, not.

Compatibilists also argue that even if determinism were true it would not be the case that there are no alternate possibilities open to us. This is typically done in two ways. The first is to provide a conditional analysis of the phrase “could have done otherwise” that is compatible with determinism. “Could have done otherwise” is equated to “could have done otherwise *if* the agent had chosen to do so.” The second way is to discuss alternative possibilities in terms of possible worlds. Though it is true that in this world I could not have done other than perform action *A*, there are other possible worlds in which I refrain from performing *A*. These possible worlds somehow account for our ability to do other than what we do.

Of course, none of these compatibilist strategies are acceptable to the incompatibilist. The compatibilist freedom to do what we want, they argue, is irrelevant if we are not also free to want what we want. And this, the freedom to want what we want, is not compatible with determinism. Additionally, conditional analysis of “could have done otherwise” are not acceptable nor is the use of other possible worlds to explain alternative possibilities. For the former, the ability to do otherwise if we had chosen to otherwise seems empty when it is added that we lack the ability to choose otherwise in a determined world. For the latter, what we are able to do in another possible world is not relevant to the discussion of free will. The problem of free will deals exclusively with our freedom or lack thereof in this world, not any other possible one.

### **1.3 Looking Ahead**

My main interest in this work is to examine the libertarian philosophy of the indeterminist. I am doing so in order to discover whether the critics are correct when they argue that indeterminism cannot aid in gaining free will and that no account of free

will or free agency (specifically indeterminist or generally libertarian) has been given. In order to do so I examine the work of Robert Kane, a leading indeterminist philosopher. In chapter two, I first describe how Kane uses indeterminism to gain freedom and provide several criticisms to his account. Most notably, that his account is not intelligible after all and that, at best, it makes compatibilist accounts of freedom and free agency more intuitive.

In the philosophical literature there are standard objections against indeterminist accounts of freedom and free agency. In chapter three, I examine four such objections as provided by Galen Strawson, Thomas Nagel, Daniel Dennett, and Richard Double. Galen Strawson questions whether indeterminism can be placed in any place that matters to questions of free will. Strawson's challenge to the indeterminist is that it is not enough to merely deny the truth of determinism, they must also provide an intelligible theory of freedom and free agency in which indeterminism plays a vital role. Nagel discusses the problem of autonomy. His worry is that the idea that we freely perform our actions is merely an illusion. We really do not *act* at all, but rather what we do is only what happens through natural and physical law. Though typically this criticism is lodged against compatibilists, it can also serve as a criticism to libertarians who are indeterminists. Like Strawson, Dennett also wonders where indeterminism can be introduced such that it makes a difference in the problem of free will. Unlike Strawson, Dennett additionally wonders about the nature of the indeterminism involved. Double lodges what I refer to as an objection from rational explanation against the indeterminist. He examines the indeterminist libertarian philosophies of Kane and Peter van Inwagen and questions how the actions that the agents perform can be considered rational. After doing so, I question whether any indeterminist account of freedom or free agency can answer these four objections and still somehow obtain libertarian freedom.

In chapter four, I disregard the various objections that have been raised against indeterminist accounts of freedom. I do so in order to question whether, even if all the previous objections are answered satisfactorily, indeterminist theories of agency do not face further problems. To do so, I follow a philosophical debate between van Inwagen on one side and John Martin Fischer and Mark Ravizza on the other. As I have mentioned previously, van Inwagen is an indeterminist libertarian philosopher. What I have not mentioned is that van Inwagen is responsible for several of the arguments that purport to show that determinism is false. Van Inwagen argues that in order to be an incompatibilist, one must rely upon a rule of reference similar to a rule that he has developed that he calls "Rule Beta." In addition to this, van Inwagen argues that the falsity of determinism does not imply that agents are actually able to do other than they do in a significant amount of cases. However, van Inwagen argues that this does not result in any appreciable decrease in the amount of actions for which an agent can be held responsible.

In the final chapter, chapter five, I conclude that indeterminist philosophers have not adequately answered the objections raised by myself and the philosophers discussed in chapter three. I additionally briefly sketch two further problems for the indeterminist. I first question whether it is appropriate to hold agents responsible for actions that flow from their character (supposing that they are responsible for having the characters they have). If we cannot do so (and I don't think there is any indisputable reason why we should), then van Inwagen would be wrong. Our inability to do other than we do in a significant amount of cases does result in an appreciable decrease in the amount of actions for which an agent can be held responsible. Second, I attempt to bolster the compatibilist position in order to show that several devices used by indeterminists (and libertarians in general) are not relevant with respect to the problem of free will. Most notably, I question the relevance of the ability to do otherwise and the necessity for indeterminism.

#### **1.4 End Notes**

<sup>1</sup> Van Inwagen, in O'Connor (1996), 219.

<sup>2</sup> Schopenhauer (1960); Honderich (1988). The relevant work by Hobbes can be found in Molesworth (1962); a relevant work by Mill is "From an Examination of Sir William Hamilton's Philosophy" and can be found in Morgenbesser and Walsh (1962), 57-69.

## 2. Robert Kane's Incompatibilism

In *The Significance of Free Will*, Robert Kane offers an incompatibilist account of freedom of the will that differs from most of his libertarian brethren. Kane divides his book into two sections. In the first, he addresses the questions of compatibility and significance – the question of whether free will is compatible with determinism and the question of why we should want to possess a free will that is incompatible with determinism. In the second part, he addresses the questions of intelligibility and existence – the question of whether sense can be made of free will that is not compatible with determinism and the question of whether such a freedom exists in the natural world. I devote the first two sections of this chapter to explaining Kane's answers to these four questions. In the third section, I argue that Kane's answer to the second and third are not adequate.

### **2.1. Compatibility and Significance**

In this section I discuss Kane's answers to the compatibility and significance questions. Kane takes free will in the traditional sense very seriously and argues that it is entailed by a condition of ultimate responsible that is not compatible with the truth of determinism. Additionally, he argues that the traditional sense of free will is a significant freedom that is worth wanting.

#### **2.1.1. Taking Free Will Seriously**

Unlike compatibilist philosophers who prefer to speak in terms of free action, Kane holds freedom of the will to be of primary import. This is so because he takes the traditional idea of the will very seriously as opposed to other modern philosophers who

use the term free will as a nod to philosophical tradition when they are actually referring to free action. For Kane, free will is “*the power of agents to be the ultimate creators (or originators) and sustainers of their own ends or purposes*” whereas free action is merely “to be unhindered in the pursuit of your purposes”(4) regardless of the ultimate origin of those purposes.

Kane’s traditional definition of free will goes hand in hand with traditional notions of moral responsibility – we hold whoever is the ultimate cause of the action responsible for the products of the action. If I intentionally push Sheila in front of a bus, I am morally responsible for her murder. If John intentionally pushes me into Sheila so that the bus will hit her, then John is responsible for her murder and my body was merely his instrument. The second case is uncontroversial and both Kane and compatibilists would agree that John, not I, is the guilty party. The first case is another story. Kane would only hold me responsible if the ultimate cause of my action rested within me. If the chain of causality can be traced outside of myself, say to my genetic history or my environment, then I am not the culprit. Though I did, say, internally form an intention and purpose to push Sheila, I could exhibit no control over whether or not the intention arose and whether or not I acted upon it. Kane would not hold me morally responsible because of the importance of free will, while a compatibilist, acting under the definition of free action, would hold me responsible so long as the action I performed was the action that I wanted to perform.

Kane traces this divergence in the modern era to the debates between Hobbes and Bramhall.<sup>1</sup> The differences between Kane and compatibilists can be seen clearly via an analysis of this debate. Hobbes took the free action position and argued that freedom of

the will as traditionally defined is unintelligible. Freedoms ordinarily desired by humans, he argued, freedom from physical restraint, coercion, compulsion, and oppression, are compatible with determinism. We are free so long as we are self-determining, and we are self-determining so long as nothing prevents us from doing what we will. This type of freedom can be possessed even though what we want or intend to do is determined by antecedent circumstances or causes.

Bramhall took the traditional free will position, arguing that the type of freedom professed by Hobbes is no freedom at all. True freedom of the will, the freedom that does matter, is not only the freedom to do what we will, but also the additional freedom for the will to determine itself. Without this freedom we are like the falling rock that is able to do whatever it wills so long as it wills to go down. Hobbes responded by pointing out a dilemma that still haunts libertarian accounts of free will.

In order for the will to have ultimate control over itself, Hobbes noted, some of its acts must be undetermined. But undetermined actions do not equate to freedom because whatever is undetermined is not controlled by anything, the will and agent included. The libertarian dilemma is one of either confusion or emptiness, the confusion of equating freedom with indeterminism or the emptiness of positing accounts of self-determination that could not be explained.<sup>2</sup> Kane (obviously) sides with Bramhall in the debate but agrees that no intelligible answer to the libertarian dilemma has been presented. Before he can attempt to do so, he must first demonstrate that there is a type of freedom worth wanting, an intelligible freedom, that compatibilist freedoms do not encompass. He must justify his libertarianism by demonstrating that it better captures our intuitions than the compatibilist position. The rest of this section will detail how he does this.

The traditional question of whether freedom is compatible with determinism is too simple for Kane because it implies that there is only one type of freedom at hand. There are actually many kinds of freedom involved in philosophic debate, as seen earlier in the distinction made between free action and free will. The question is best formed thusly: “Is freedom *in every significant sense worth wanting* compatible with determinism?”(14). For Kane to succeed he does not need to show that *no* significant type of freedom worth wanting is compatible with determinism - he admits that many of them are compatible. All he must demonstrate is that there is at least *one* significant type of freedom that is not compatible with determinism to show that the compatibilist position is untenable. This freedom is, of course, freedom of the will.

A traditional argument against compatibilism rests upon the idea of alternate possibilities as a necessary condition of freedom of the will. In the next sub-section, I discuss Kane’s treatment of alternate possibilities and show why he rightly claims that it alone is not sufficient to show that compatibilism is lacking.

### **2.1.2 Could Have Done Otherwise – Alternate Possibilities (AP)**

Kane provides an analysis for determining whether or not an action is “up to an agent” in the sense necessary for ascription of free will as follows:

(AP) The agent has *alternate possibilities* (or can do otherwise) with respect to A (an action) at time t in the sense that, at t, the agent *can* (has the *power* or *ability to*) do A and *can* (has the *power* or *ability to*) do *otherwise*. (33)

This is central to the notion that an agent should not be held responsible for an action if he was unable to do other than he did. Under normal circumstances we are held responsible for the results of our actions because we seem to have a choice concerning which actions to undertake. Though Harry has chosen to see *Black Hawk Down* at the

theater it is also true that he could have chosen not to go in favor of riding his bicycle by a lake. He could have done either and because of this he is responsible for that action he does actually do.<sup>3</sup> However, if Harry was hypnotized in such a way that whenever he is faced with a choice of going to the theater or going for a bicycle ride he will always choose the bicycle ride, then it is not true, in this limited situation, that Harry is able to do other than go to the theater. Not being able to do otherwise undermines freedom and responsibility and is a threat to compatibilism. If determinism is true, then it would seem that it is never the case that an agent could have done other than what he does in fact do.

Compatibilists have mainly attempted to refute this in two ways. First, they have argued that it is not necessarily the case that we do not hold agents responsible in cases where it is agreed that they could not have done otherwise. Second, they have argued that even if determinism were true there are analyses of “can” and “could” with which it is true that agents could have done otherwise and can do otherwise. Kane discusses Dennett’s Martin Luther example and Frankfurt’s Black and Jones example in reference to the former.<sup>4</sup>

Dennett argues that when Martin Luther broke with the Church of Rome and stated, “Here I stand. I can do no other,” it was true that Luther could not have done otherwise yet we still hold Luther’s act as one for which he can be held accountable. With his statement, Luther was taking full responsibility for his action rather than avoiding responsibility. If this were so, a condition like AP would be necessary neither for moral responsibility nor free will in any sense worth wanting. Rather than caring about whether an agent could have done otherwise when assigning moral responsibility, Dennett argues, we consider whether the consequences that flow from the action are good

or bad and also whether or not praising or blaming the agent for the action can modify the agent's and other agents' future actions. If Sam steals a car and, like Luther, his character was such that he could not have done other than do as he did, we would hold Sam as morally blameworthy and punish him because it would make Sam and others like him less likely to perform unacceptable acts in the future. This is so because agents generally do not wish to be on the receiving end of such punishment.

Kane cites Dworkin<sup>5</sup> as providing an adequate refutation of this last point. Dworkin notes that moral ascription of this sort are inadequate because they are forward looking and do not take into account whether or not a person deserves to be praised or blamed for his action. In order to determine whether an agent is blameworthy or praiseworthy, we must look to the past and not the future and examine how the agent came to be the type of person that they are. In the case of Luther, moral accountability depends upon whether Luther is responsible for being the sort of person that he was at the time, not upon whether the future effects of holding him responsible would be favorable.

Kane agrees wholeheartedly with the last of this. In order for an agent to be held morally responsible, it is not necessary that they could have been able to do otherwise in every single instance so long as the agent is ultimately responsible for his inability to do otherwise. For Luther, this would be so because at some point in Luther's past he could have done otherwise, he could have chosen to keep his faith private, and thus would not be in the position described. At some point he could have done otherwise.

In Frankfurt's Black and Jones example, Black is an evil neurosurgeon with direct control over Jones' brain and has intimate knowledge of Jones' proclivities. Black wants Jones to perform a certain act, say, voting for Bush in the primary election. Black knows

Jones well enough to predict which way he will go. If things are going such that it looks like Jones will vote for Gore, Black will press a button which overrides Jones' will and forces him to vote for Bush. If it looks like Jones is going to vote for Bush, however, Black will do nothing and Jones will follow his own will and cast his vote. In this latter alternative, it appears that Jones can be held responsible for his vote for Bush even if, as the first alternative shows, Jones could not have done other than vote for Bush. If he were leaning towards Gore, Black would have known and forced the Bush vote. Kane uses examples like this as an argument that AP does not sufficiently show that compatibilism is false. This is because, as shown in the Luther example, Kane does agree that we can be held responsible in cases where we could not have done otherwise. Similarly in the Black and Jones example, Jones can be held responsible for his vote for Bush in the second alternative even if he could not have done otherwise. However, in either alternative of the Black and Jones example, more investigation is required to determine whether Jones should be held responsible. In the former case, we would not hold Jones responsible after examining his past because we could see that he was going to vote for Gore until Black interfered. In the latter case, we may or may not hold Jones responsible for his vote depending upon whether or not Jones was responsible for having the type of character that necessitated a vote for Bush.

What both cases show for Kane is that AP is not sufficient reason to be an incompatibilist, but AP does point to something that is sufficient, ultimate responsibility, which will be discussed in the next sub-section.

The second way that compatibilists have argued against AP is arguing that even if determinism were true there are analyses of "can" and "could" with which it is true that

agents could have done otherwise and can do otherwise. Kane discusses these in terms of van Inwagen's Consequence Argument, which states:

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 present acts) are not up to us.<sup>6</sup>

If this is correct, then if we were able to do otherwise it is in our power to either change the past or falsify a law of nature. Since we can do neither, then it must be true that if determinism is true we are unable to do otherwise. Kane claims that this argument does succeed for free will, unless the compatibilist can provide a compatibilist account of *can* or *power* that succeeds.

Kane discusses several attempts by compatibilists to show that the argument does not hold. A traditional analysis of "could have done otherwise" is the conditional analysis. Within the "could" is a buried conditional – "could have done otherwise" becomes "could have done otherwise if the agent had so chosen". This analysis is compatible with determinism because it can be true that an agent could have done otherwise if the agent had so chosen while it also being determined that the agent could not have so chosen. Van Inwagen considers and rejects this analysis because from it we can deduce that an agent could change the past or break a law of nature if the agent so chose, and that seems false if not simply very counterintuitive.

However, David Lewis has argued that this can be made sense of if a "weak sense" of "being able to render a proposition false" is employed.<sup>7</sup> I can render false a proposition in the "strong sense" just in case "I was able to do something such that, if I did it, the proposition would have been falsified, either by my act itself or by some event caused by my act."<sup>8</sup> I can render false a proposition in the "weak sense" just in case "I

was able to do something such that, if I did it, the proposition would have been falsified (though not necessarily by my act, or by any event caused by my act).”<sup>9</sup> The weak sense only entails that if an agent had acted otherwise, then a law of nature would have been different, not that the agent caused the law of nature to be different. In this weak sense, it is true that we can render a law of nature false. [But it is not clear to me that rendering a law of nature false actually amounts to anything. The laws of nature are immutable. If a law of nature were broken in the strong sense, we would perhaps say that we were mistaken about the status of the law in the first place. For example, if a particle is discovered that travels faster than the speed of light, we would not say that a law of nature had been broken but rather that Einstein was wrong about what the laws were. The case of the weak sense is not comparable to the strong sense. What passes for breaking a law of nature in Lewis’ weak sense is merely the claim that the laws of nature could have been different and, hence, could have necessitated a different action than the action it did in fact necessitate. Rather than claiming that an individual breaks a law of nature in any sense, it would be more appropriate to say that the law of nature breaks the individual.<sup>10</sup>]

Kane notes that conditional analyses of “could have done otherwise” have also come under attack by J. L. Austin and Roderick Chisholm.<sup>11</sup> Austin argues that the statement:

(C) You could have done otherwise.

cannot be equivalent to the statements:

(CI) You *could* have done otherwise, *if* you had willed or chosen or wanted to do otherwise.

or

(WI) You *would* have done otherwise, *if* you had willed or chosen or wanted to do otherwise.

CI cannot be correct because it makes the existence of a power or ability to do something dependent upon an agent's willing or choosing to exercise the power or ability. It is absurd to say that I do not have powers that I do not exercise, for surely I have the power to jump off the Empire State Building even if I never choose to do so. Additionally, WI cannot be adequate because it implies that we can succeed in doing whatever we set out to do. The example Austin gives is of a three-foot putt. Making the putt is certainly within his power, but that does not mean that he is guaranteed of making it should he attempt it. After missing the putt it is true that Austin could have chosen to make the putt, but that does not entail that he *would* have made the putt.

Chisholm argues that WI and CI do not adequately capture the truth of C unless a further condition is added:

(C') You could *also* have *willed* or *chosen* otherwise.

However, C' introduces the troublesome "could" again which calls for another conditional analysis:

(WI') You would have chosen otherwise, if you had *willed to choose* otherwise.

that, in turn, requires another C' type condition stating that one *could* have willed to choose otherwise, and so on. Chisholm points out that this would regress infinitely with each subsequent WI requiring a C' and each C' requiring a further WI'. The regress would not allow for the elimination of "could".

These differences have resulted in an impasse over the importance of AP. Kane thinks that neither the compatibilist nor the incompatibilist has provided a case

convincing enough to the other. He argues that this is so because a condition like AP is not sufficient to eliminate the compatibilist position. More is required, in this case the joint condition of ultimate responsibility (UR).

### **2.1.3. Ultimate Responsibility (UR, U and R)**

Kane argues that AP alone is not enough to win the day for incompatibilists - “focusing on the power to do otherwise and alternative possibilities alone is just too thin a basis on which to rest the case for incompatibilism” (59). In addition to AP, and what in fact AP and most debates concerning free will point towards, is condition UR which is made up of two subconditions, U and R:

(UR) An agent is *ultimately responsible* for some (event or state) E’s occurring only if (R) the agent is personally responsible for E’s occurring in a sense which entails that something the agent voluntarily (or willingly) did or omitted, and for which the agent could have voluntarily done otherwise, either was, or causally contributed to, E’s occurrence and made a difference to whether or not E occurred; and (U) for every X and Y (where X and Y represent occurrences of events and/or states) if the agent is personally responsible for X, and if Y is an *arche* (or sufficient ground or cause or explanation) for X, then the agent must also be personally responsible for Y. (35)

Kane notes that the first subcondition, R, can be given a compatibilist reading with conditional analyses of “could” in “could have voluntarily done otherwise.” Because of this, as shown in the last sub-section, R alone is not enough reason to be an incompatibilist. It is in the second, backtracking subcondition, U, where incompatibilism is shown to be a necessity.

Consider Paul, a rampant womanizer, and Joan, his latest victim. Paul tells Joan whatever she would like to hear (lies, of course) in order that he may take her to his bed. According to R, we can hold Paul responsible if he could have voluntarily done other than what he did. Given a compatibilist spin, this becomes if he could have voluntarily

done other than what he did if he had so chosen to do so. Kane may not be happy with conditional analyses of “could”, but he will not raise a fuss because it is in U where true responsibility lies. It is not enough that Paul could have voluntarily done other than what he did. He must also have been responsible for whatever would have allowed him to do so. Under the compatibilist reading of “could”, Paul plays no causal role in his possibly doing otherwise, rather, his doing otherwise would result from the past being different or the changing of a law of nature. Neither option is incredibly likely, even given Lewis’ weak sense of being able to make a proposition false.

To simplify, let’s say that there was a single action *A* in Paul’s past which led him to become a rampant womanizer. According to U, Paul is only responsible for his current action provided he is also responsible for *A*. But it does not seem possible for Paul to be responsible for *A* if determinism holds because *A* would have a cause, *B*, of its own which Paul must have been responsible for, and *B* would have cause *C*, etc., until it regresses to a point before Paul existed. For Kane (under U), the causal chain must stop at a point where Paul is still capable of being responsible and is in fact responsible for the stoppage. *A* would then have to be not determined by prior events yet somehow be caused by Paul. Kane refers to an action of this type as a self-forming action (SFA) or self-forming willing (SFW). Kane defines an SFA as:

SFAs are the undetermined, regress-stopping voluntary actions (or refrainings) in the life histories of agents that are required if U is to be satisfied, and for which the agent is personally responsible in the sense of R. The agents must therefore be responsible for them directly and not by virtue of being responsible for other, earlier actions (as would be required if they were not regress stopping). This means that, for SFAs, the “something the agents could have voluntarily done (or omitted) that would have made a difference in whether or not they occurred” is simply *doing otherwise*, rather than doing something *else* that would have causally contributed to their not occurring. (75)

In the case of Paul, action *A* was an SFA and, because of this, Paul is ultimately responsible for both *A* and the subsequent action of seducing Joan.

#### **2.1.4. The Significance of Free Will**

Having established that the type of freedom in question, freedom of the will, is not compatible with determinism by UR, Kane turns his attention to providing reasons for accepting UR. To do so, he discusses the concept of sole authorship or underived origination. This concept is considered at one time or another by both compatibilists and incompatibilists, to be embraced by the latter and rejected by the former. This concept holds the source of action to be the agent or self and not something outside of the agent. The causes of our actions would be traceable back to a SFA of which the agent is the sole author and underived originator. It is this type of free will that ordinary persons believe they want when they want free will.

This type of freedom has typically been seen to be worth wanting because it is necessary for other goods that are generally desired and are worth wanting. Among these other goods are genuine creativity, self-legislation, true desert for one's achievements, dignity, moral responsibility, etc. (80). Kane describes what he calls the dialectic of underived origination or sole authorship which begins with incompatibilists arguing that the goods mentioned are not compatible with determinism. For example, the truth of determinism would entail that *Starry Night* is no more an achievement of Van Gogh than it is of me. The creation of the work was inevitable and there is nothing in the work that originated within Van Gogh but rather was caused by events prior to his birth.

The second step of the dialectic is the compatibilist response. The compatibilist argues that the goods mentioned above are possible without UR. Even if determinism

were true, *Starry Night* would still be an original work that was not created before Van Gogh and was only possible through Van Gogh. According to the compatibilist, the incompatibilist objection is question begging with respect to the falsity of determinism – they describe these goods in such a way that they cannot be compatible with determinism. It is, therefore, not surprising that the goods are not available if determinism were true. However, there are other accounts of the goods in question that are neutral with respect to the truth of determinism, and it is these the incompatibilist must draw from for their argument to hold. The incompatibilist responds that these other accounts do not capture what is worth wanting in creativity and the other goods. There is a more exalted sense in which we want to be able to create. To this the compatibilists respond that the incompatibilist begs the question and an impasse is reached again. Kane recognizes that most free will debates do not get beyond the impasse that results from the dialectic of origination. He argues that we must dig deeper into the conflicting intuitions behind the impasse.

Free will is a metaphysical issue in that it deals with the ultimate source or explanation of responsible human actions. What results from examining the deeper metaphysical problem of free will is not the dialectic of origination, but rather what Kane calls the “dialectic of selfhood.” In this dialectic, Kane tells a story of an infant who in the midst of interacting with the world learns that she can control certain things in her environment, like her hand, and not control others. She learns that the hand is part of her and that she can control it via an act of will. In this way the infant learns to separate herself from the world as an independent causal agent. As the infant grows older, she feels the need for approbation – appreciation and acknowledgement for what she does.

Kane suggest that this desire for approbation is part of a fundamental need to affirm her selfhood as an independent being that is a source of activity. It is this more fundamental need that serves as the basis of the goods mentioned in the dialectic of origination (creativity, autonomy, etc.).<sup>12</sup>

The awareness that she is a part of and causally influences the world brings about a spiritual crisis. The crisis takes the form of the worry that just as the world is causally influenced by her, so she is in turn causally influenced by the world. This is the fear that she is not separate from the world at all but merely a part of it. This is the traditional fear that we possess no free will but are mere physical beings to whom freedom is but an illusion.

Kane considers two possible reactions to this spiritual crisis. The first is that she insists that she is not part of the physical world at all but rather can still causally act upon it. This is Cartesian dualism and Kane finds this reaction too crude. The second, a less crude reaction than the first, does not place the self completely outside the world. She is part of the world and is influenced by it but she somehow has the final say on which way she is influenced. Kane uses as an analogy the membrane of a cell that allows in that which is useful to the cell and keeps out that which is harmful. In this way the agent can imagine herself as a sophisticated being with the selective power to choose how she affects and is affected by the world. Inside her “membrane”, she is able to find refuge from the spiritual crisis.<sup>13</sup>

This second reaction can only be a temporary solution for the agent for she will surely realize that she is neither completely in control of nor completely aware of all of the outside influences. Here the pervasive threat of determinism comes completely to the

fore. She cannot be sure that the choices that she makes within her “membrane” are not determined by her nature and are therefore not in her control. Kane suggests that we view the thread of determinism not as an isolated phenomenon but rather as a stage in the dialectic of selfhood. At each stage of the dialectic, she tries to preserve the idea that she is an independent source of activity. From this stage she is propelled to an expression of UR. A conviction that though many of her choices may be determined, it cannot be so for all of her choices. In this way Kane sees free will as a “*higher stage response to the dialectic of selfhood*” that “emerges as an issue when we realize how profoundly the world influences us in ways of which we are unaware” (96).<sup>14</sup>

Kane provides the example of Alan the artist to demonstrate another reason we find free will to be significant: objective worth. Kane asks us to consider two worlds. In both, Alan’s paintings have not found the success that he would have liked. In the first world, a rich friend of Alan’s secretly arranges to buy several of Alan’s works through agents acting on his behalf. In the second world, the purchasers of Alan’s works do so because they genuinely find them admirable. In both worlds, Alan dies happily believing that he is successful artist, but it is only actually true in the second world. Though both worlds are subjectively identical for Alan, we do think that there is a reason to choose the second world over the first. For Kane, this reason is that the objective worth of our actions does matter. The fact that we do consider the objective worth important shows that we are not merely concerned with how things appear to us (whether it merely appears to us that we are free), but rather how things actually are (whether we are actually free or not). If we did not find free will significant then we would not hold

things like objective worth important. The fact that we do shows that we hold free will as significant.

It is important to note that Kane does not offer the dialectic of selfhood (nor the importance of objective worth) as some sort of proof or argument that freedom is not compatible with determinism. He has already established that the freedom he is concerned with, freedom of the will, is not compatible with determinism via his discussions surrounding UR. What he is attempting here is to show both *that* this freedom is significant and show *why* it is deemed so. This is the role of the dialectic of selfhood. Freedom of the will may turn out to be something unintelligible, but whatever it is it will be something that agents desire and hold as important.

Having established that free will is significant, Kane next attempts to develop a conception of free will that is intelligible.

## **2.2. Intelligibility and Existence**

In this section I discuss Kane's response to the intelligibility and existence questions. He attempts to answer the former by appealing to plural rationality and indeterminate efforts of will. The latter he answers by utilizing quantum indeterminacy, chaos theory, and folk psychology.

### **2.2.1. The Free Agency Principle**

Traditional compatibilist attacks against libertarians have focused on the unintelligibility of their position – the mysteriousness that goes with the emptiness of accepting the second horn of the libertarian dilemma by positing accounts of libertarian agency that cannot be adequately explained. Kane hopes to make libertarianism at least on par with compatibilism by not allowing the libertarian to call on any special entities or

special forms of causation to explain free will. Kane only allows one tool that the compatibilist is not allowed to utilize – indeterminism.<sup>15</sup> To do so, Kane formulizes “The Free Agency Principle” (FAP). Under this principle, the incompatibilist is allowed “that some of the events or processes in libertarian free agency will be indeterminate or undetermined events or processes”. However, these events or processes cannot be explained by an appeal to “categories or kinds of entities that are not also needed by non-libertarian (compatibilist or determinist) accounts of free agency” (116). Out go Kantian noumenal selves, Cartesian Egos, and special types of agent causation. These libertarian strategies had their hearts in the right place but must be set aside in order for the incompatibilist position to put itself on the same ground as compatibilists with regards to their relation to modern science. If Kane can perform such a task, he will have struck a marked blow for libertarian philosophy. He attempts to do so by appealing to plural rationality, quantum indeterminacy, chaos theory, and folk psychology.

### **2.2.2. Plural Rationality – The Divided Will**

A compatibilist criticism against the use of indeterminism to explain free will is one-way rationality. Suppose Lance ventures to a sporting goods store to purchase a mountain bicycle so that he can bike along several forest trails while on vacation. When he arrives, he notices that there are two types of bikes available: mountain bikes and street bikes. Once there, Lance can choose to purchase a mountain bike or he could choose to purchase a street bike. The street bike, with its skinny tires and low durability wouldn't suit Lance's purposes at all. Which type of bike will Lance actually purchase? From the rational point of view, he really has only one choice – the mountain bike. The decision to purchase the mountain bike would be a rational decision, the only rational decision available. If Lance's choice was

indeterminate and resulted in the purchase of the street bike, we could consider his purchase foolhardy and irrational. This is what is meant by one-way rationality. When faced with a choice, only one option is the most rational one and, hence, is the only rational option to choose. Any other choice would be at the very least less rational than it.

Kane argues that libertarians must give up one-way rationality if they hope to achieve an account of indeterministic freedom and agency that is intelligible. This is so because the libertarian must make allowances for the ability to have done otherwise. This ability amounts to very little if it is only the ability to act irrationally. Because of this, libertarians should accept plural rationality. Under Kane's account, Lance's decision to buy the street bike would never occur because Lance has no reason to buy the street bike. As we shall see later, the conflict needed for the indeterminacy to arise does not occur. A more apt example would be one of Greg who also wishes to buy a bike and must choose between a mountain bike and a street bike. Like Lance, Greg also wishes to ride along forest trails. Unlike Lance, Greg also desires (say, to a lesser extent) to race against other street bike riders in a race. In this example, it would be rational for Greg to choose the mountain bike because it is what he most wants to do. However, Greg does have a desire (and, hence, a reason) to purchase a street bike. If he goes home with a street bike, it will not be an irrational decision because, unlike Lance, Greg had reasons for purchasing the street bike. In a case like this the will is best thought of as divided. Before the decision is actually made, Greg's divided will supports the selection of either bike.

Kane has similar arguments in favor of plural accounts of voluntariness and control. Given the setup of the example, it would seem odd to say that had Lance chosen to purchase the street bike he would have done so voluntarily and it would have been a

choice of which he was in control. We wouldn't say that because he doesn't have any desire to purchase the street bike. The case of Greg is different and provides an example of plural voluntariness and plural control. Because he has reasons for choosing either bike, either resulting choice would be voluntary and in Greg's control. This will be discussed more later. Now I turn my attention to the role that indeterminacy and chaos theory play in Kane's account of agency.

### **2.2.3. Indeterminacy and Chaos Theory**

Though universal determinism has been in retreat in the physical sciences due to the advance of quantum physics, it has not led to an increase in indeterministic theories of freedom. This can be explained because of trends within sciences other than physics, most notably biology and the social sciences, which have convinced many that more and more of our behavior is determined by causes that are not known to us and beyond our control. Additionally, indeterminacy at the micro level does not seem to have any obvious indeterminate effect on the macro level, which include larger physical systems such as the human brain and body. Compatibilists have taken this line and further argued that even if indeterminacy were to have macro effects it would not help the indeterminist's position. Action that is indetermined is not action but simple motion for which an agent cannot be held responsible. In appealing to quantum indeterminacy, Kane must give both an explanation for how micro indeterminacy can cause macro indeterminacy and explain how it results in an action for which the agent can be held responsible. Kane explains the latter via a materialistic view of the self and folk psychology (which I go into in the next sub-section) and explains the former via chaos theory.

Chaos theory involves the notion of sensitivity to initial conditions. Very minute changes in the initial conditions grow exponentially and result in very large differences in the final outcome. The apparently insignificant fluttering of a butterfly's wings in China, for example, can via chaotic effects result in rain falling on Central Park. Similarly, the seemingly negligible indeterminacies at the quantum level can, via the perturbation amplification of a chaotic system, result in indeterminacy at the macro level. Kane cites current work in neurophysiology that indicates that neural networks can express chaotic effects. I now turn to showing what role indeterminism plays in an agent's decision making process.

For Kane, the opportunity to perform a "self-forming action" or a "self-forming willing" for which the agent is ultimately responsible occurs when a divided will arrives at a choice that must be made among non-compossible alternatives. A typical example is an agent whose will is divided between following a moral course of action and prudential course of action but cannot do both. A shopkeeper must decide whether or not to overcharge her customers. If she does, she will (conceivably) earn more. However, if she does overcharge she will have acted immorally – against her own morality. Every agent possesses this sort of divided will to some extent that results in two competing desires, the desire to be prudent and the desire to act morally. The shopkeeper, like most of us, would rather be moral but it takes some amount of effort of will to resist the desire to act self-interestedly. In such situations of conflict, what the shopkeeper will do is uncertain – even to the shopkeeper. This is because in cases of struggle between a divided will she cannot know before hand which side will win out. It is in this uncertainty that Kane places the

indeterminism and “(t)he uncertainty and inner tension that agents feel at such moments are reflected in the indeterminacy of their neural processes” (130).

After the choice is made, she will (because of plural rationality) be able to look backwards and provide reasons for making that choice (prudential reasons on the one hand, moral reasons on the other). However, what has actually happened is that anxiety over non-compossible choices has had a chaotic effect in her on activity on the quantum level. This results in the opening of a window of indeterminacy at the macro level that enabled her to make a “self-forming action” or “self-forming willing” for which she is responsible. The complex process involved in the indeterminacy is felt phenomenologically as an effort of her will. Or, rather, the indeterminate process in the brain is a physical realization of her effort of will.

To further explain how the action is an action for which the shopkeeper can be held responsible, I need to discuss Kane’s materialistic view of the self. He equates the self with a self-network that is a neural net. He follows Owen Flanagan by considering the self as a model contained in the brain. It is this model (which plans, aspires, etc.) that Kane identifies as the self-network. He argues that the unity of the self-network can be found “in the dynamical properties of neural circuits and connections that make such synchronous patterns of neural firings possible” (140). The neural events that correspond to our efforts and choices are, in this theory of agency, overlaid by wave patterns which unify the self-network, “so that the wave patterns and the effort or choice events are coupled, causally influencing and interacting with each other” (ibid.). These “superimposed patterns of oscillations” would be contributing causes to choice by pushing one competing “reason-network” to the forefront. In the case of the shopkeeper, one reason-network would support

charging fair prices and another would support overcharging. The choice ultimately made is indeterminate in a sense (because the effort of will is influenced by quantum indeterminacy) yet it is still a choice made by the agent for which the agent, according to Kane, can be held responsible. This further explains how character is formed according to Kane. The indeterminate process results in a decision that in turn affects the state of the self-network. In the case of Martin Luther, his earlier actions in life helped form his self-network such that his later decision was determined.

Indeterminacy acts to maintain the ultimate responsibility of the agent by breaking the causal chain (that results in an action) within the agent herself. We cannot defer to conditions that held before the shopkeeper existed to explain why she acted morally rather than immorally, the causal chain of explanation ends inside the agent via her indeterministic effort of will. Her action can be explained (by either moral or self-interested reasons), but the cause of the resulting action is ultimately the agent. In this way, Kane satisfies condition UR.

#### **2.2.4. Folk Psychology**

From a purely physical point of view, it is hard to accept that it is the shopkeeper that does anything. Kane notes as much by stating that “when neuroscientists described it (her action) in physico-chemical terms, all they would get are indeterministic chaotic processes with probabilistic outcomes” (147). However, Kane argues, the scientific perspective is not the only perspective from which to assess an agent’s action. There is also what Kane refers to as the *phenomenological perspective* from which, *experientially considered*, the physical process *is* the agent’s choice. For this reason, Kane argues that one can’t be an eliminative materialist with regard to human action - it is in virtue of the folk

psychological descriptions that we are able to ascertain that we are free. The physical description cannot be the *only* description available. To do so would be to write free will out of the world picture along with other valued things such as consciousness, purpose, and mental action in general (ibid.).

It seems counterintuitive for us to believe that consciousness is a physical process, but this is a problem that Kane argues is shared by any (materialist) account of free agency, compatibilist or incompatibilist. “It is no less mysterious how neural firings in the brain could be conscious mental events if they are *determined* than if they are undetermined, or if they involved undetermined chaotic processes than if they do not” (148).

Indeterminism and folk psychology play vital roles in Kane’s theory. Without indeterminism, an agent cannot be ultimately responsible for her action. Without folk psychology, his materialist conception of the self and human action would not allow that the undetermined choice was something the agent did as opposed to something that merely happened.

### **2.3. Criticisms**

Kane has done an admirable job of creating a libertarian account of free will that is at the very least an improvement over traditional libertarian accounts of agency that are eliminated by the Free Agency Principle. However, all is not well in his libertarian paradise. In this section, I outline several objections to Kane’s libertarian philosophy. The first sub-section will contain criticisms of Kane’s argument for the significance of libertarian free will. The second sub-section contains criticisms of his use of indeterminacy to gain freedom of the will.

### 2.3.1. Significance

The first set of criticisms concern Kane's attempt to answer the significance question via the dialectic of selfhood. Recall that Kane considers the worry of determinism (and hence any compatibilist position) merely a stage in the dialectic of selfhood, one that is surpassed by the higher stage of becoming a believer in freedom of the will. Kane has not provided sufficient justification for his ordering of the stages in this manner. His chosen stopping point, the stage of free will, is arbitrary. His suggestion is that it is common for agents, through the course of their lives, to engage in the dialectic. If this is so, surely then some of those who engage in the dialectic of selfhood are the very compatibilist philosophers against whom Kane is arguing.

But then Kane would have to provide a plausible explanation for what went wrong in their case, an explanation that maintains the supremacy of his final stage. Arguing that the compatibilists are stuck on a lower stage would not work. By virtue of what is the Kane's free will stage higher than the compatibilist stage? The compatibilists could even admit that Kane's dialectic is well formed but incomplete, lacking an even higher stage at which point the agent becomes disillusioned with libertarian freedom because of, say, its unintelligibility, and reaches a still higher stage of compatibilism. Perhaps compatibilism does not even come into the picture at the earlier stage that is dominated by the threat of determinism but rather only arises after the agent becomes dissatisfied with libertarian free will. Kane's story can thus be read as a just so story.<sup>16</sup> It is manufactured to back up Kane's philosophy and certainly sounds plausible but there are alternative stories which match the evidence yet do not go hand in hand with his theory, especially concerning where Kane chooses to end his dialectic.

Also problematic is Kane's use of the example of Alan the artist. Recall that in the example Alan believes himself to be a respected artist in two different worlds, but he is mistaken about this in the first world and correct about it in the second. If given a choice, Alan would choose to live in the second world (and if we were in Alan's shoes, we'd choose the same). Kane uses this to show that subjective worth is not all that matters to us. We want our subjective experiences to match with objective reality. It is not enough that Alan believe that he is a successful artist, he must objectively *be* a successful artist.

Analogously, if we were given the choice of living in one of two worlds, the first a determined one (Compatibilist World, or CW) and the second a world in which libertarian free will functions (Libertarian World, or LW), we would choose the second world. Because we would choose the second world over the first, Kane argues, we consider freedom of the will something significant and worth wanting. I argue that compatibilists can readily admit that they would prefer to live in the libertarian world without admitting that libertarian free will is significant. Imagine a third world to compete with CW and LW. In this third world, in addition to having libertarian free will we also possess the freedom to defy the law of gravity. I'll call this world Flying Libertarian World, or FLW. If asked to choose among CW, LW and FLW, surely both compatibilists and incompatibilists would choose to live in FLW (the non-acrophobic ones at any rate). Consider a fourth world that is identical to CW except that we have the ability of flight as in FLW (call this one CFW). If asked to choose between CFW and LW, there is no guarantee that LW would be chosen more often. Suppose CFW is preferable (if flying is not attractive enough, I can create other compatibilist worlds –

worlds where the past can be altered, worlds where we are all gods, etc. – complicated worlds and even, pardon the phrase, possibly impossible ones at that). By Kane’s reasoning, that would indicate that freedom to fly is a significant freedom and one worth wanting, a freedom even more significant than libertarian freedom. My point here is not to show that the freedom to break the law of gravity is more significant than libertarian free will. I only hope to show that the process Kane uses to determine whether a freedom is significant or not is unreliable. This is so because the process that he employs will allow for unintelligible freedoms or freedoms irrelevant to the problem of free will to become significant freedoms. This problem can more clearly be seen when we consider the connection between the questions of significance and intelligibility.

One final criticism also centers on Kane’s criteria for significant freedoms. Using Kane’s criteria the compatibilist would no doubt find libertarian free will significant. However, for compatibilists the question of significance is closely tied with the intelligibility of freedom at hand in a way that Kane has not accounted for. What matters most for the question of significance is whether the freedom in question could conceivably exist. Kane’s argument for significance, if it works, will only do so provided he can defend an account of libertarian free will that is intelligible. Until he does so, LW doesn’t become available as an option for choice. I now turn to criticisms that indicate that he has not succeeded in doing so.

### **2.3.2. Indeterminism and Folk Psychology**

In this sub-section, I launch two main criticisms against Kane's use of indeterminacy to gain freedom of the will. The first concerns the problem of moral luck and Kane’s

response to it. The second concerns the role folk psychology plays in Kane's libertarian philosophy and is developed into the form of a dilemma for Kane.

The problem of moral luck is especially relevant to Kane's libertarianism.<sup>17</sup> This can be seen more easily in light of the self-forming actions or willings (again, SFAs and SFWs). An SFW results not only in an action for which the agent can be held responsible, but also serves to shape an agent's character such that he will be more likely to act in a similar manner in the future. Consider John, a college student who is considering cheating on a chemistry exam because he has not studied properly. John must choose to either act morally and fail the exam, or act (arguably) prudentially and cheat on the exam to avoid the consequences of failing. Further suppose that John's character up to this point could be numerically measured and represented as a ratio representing the strength of his desire to perform either action on a scale of 100. In this case, John's character can be represented as the ratio 55:45, with the larger number designating the stronger desire.<sup>18</sup> In this case, the stronger desire is to cheat (55) and the weaker to act morally (45). Given that the two alternatives are non-compossible, John agonizes sufficiently enough that he is able to perform an SFA.

Usually agents strive against prudential choices in favor of moral ones, but not in this case. John actually desires to cheat more and, if he does not end up cheating, it will be because his effort to decide to cheat (his effort of will) failed as a result of the indeterminate process that Kane describes. John's SFA results in the moral choice. This result also has an effect on John's character such that he is more likely to perform moral actions in the future (Kane argues that the resulting choice in an SFA in turn affects the organization of the self-network in this manner).

Though he did not cheat, John managed to avoid failing (the exam was not as difficult as he feared). However, he has not learned his lesson. The very next week he has a Continental Philosophy exam that he has not studied for. Again, he is faced with the same choice – be moral and fail or be prudential and cheat. His previous choice has affected him in such a manner that he stills prefers to cheat over failing, but instead of favoring it in a ratio of 55/45 it is now 51/49.

Again imagine that the SFA results in moral choice that changes his character such that if he were to be placed in a similar situation again he would now desire to be moral *more* than to cheat by a ratio of 51/49 (51 representing the moral desire, 49 the desire to cheat). This scenario can be played out again and again, each time resulting in John making the moral choice and increasing his future chances of making more moral choices. At some point, John's character will be such that he will not be faced with a dilemma when placed in a similar situation. His act moral/cheat ratio would (conceivably) be 100:0. Because he no longer desires to cheat, he will no longer face the anxiety that results in the indeterminate SFA. These events have occurred in possible world number one (PW1 – the John in PW1 will now be referred to as John1).

Now, consider possible world number two (PW2) that is identical to PW1 up to the point where the first SFA occurs in the previous example. The John in PW2, call him John2, faces the same dilemma with the exact same character makeup. However, the indeterminate SFA ends with the choice to cheat instead of act morally. This SFA results in a change in John2's character ratio such that he will now favor cheating by a ratio of 60/40. Just like John1, John2 faces similar dilemmas in the future and each time happens to choose against acting morally until his character ratio is 100:0 in favor of cheating. At this point,

again, there is no dilemma. His character would determine that he cheat in similar situations.

John1 in PW1 has become a better, more moral person than John2 in PW2. But to what does John1 owe his good character? It is hard to see why we should praise John1 for not cheating and blame John2 for cheating when the only difference between them was that John1 was lucky enough to have SFA's that resulted in moral actions and John2 was not. The difference is that John2 had a successful effort of will and John1 did not (recall that each John originally possessed a stronger desire to perform the immoral act – it was John1's failure to perform the action he most desired that led to the moral action).

According to Kane, the effort of will is an indeterminate process. But because of this we can neither praise an agent for having a successful effort of will when trying to act morally nor can we blame an agent who fails such an attempt because he did not try hard enough. Whether he tried hard enough or not was simply not up to the agent – it was indetermined.

Kane addresses a similar criticism in chapter 10 as made by Bruce Waller.<sup>19</sup> He responds in two ways. First, he argues that though the effort of will is indeterminate, whether or not it is successful is not a matter of luck. Rather, whatever the result of the effort, the choice will be one that the agent voluntarily made. This is so both because he has reasons for performing either option and because he is responsible (via previous SFAs) for the limited options available to him. The indeterminate process will not result in the agent performing some wildly unpredictable action such as screaming gibberish and performing cartwheels. John's wrestling over cheating or not will have one of only two consequences –

he will cheat or he will not cheat. Since he is responsible for those being the only two options, it is not a matter of luck which option he settles on.

This response fails to address the criticism directly. In the case of John1 and John2, at the time the agents had identical makeup, Kane's answer does not yet give us any reason to consider that John1's resulting good character and John2's resulting bad character were not the result of moral luck. Though it might not be pure luck and only luck that decides it, there cannot be a doubt that fortune plays a hand.

The second response to this objection by Kane is to argue that the example is flawed because it assumes that the pasts of the two agents, John1 and John2, are exactly the same. This is because exact sameness is not defined with indeterminate efforts. As Kane states rather strongly:

If the efforts are indeterminate, one cannot say the efforts had exactly the same strength, or that one was exactly greater or less great than the other. That is what indeterminacy amounts to. So one cannot say of two agents that they had exactly the same pasts and made exactly the same efforts and one got lucky while the other did not. Nor can one imagine the same agent in two possible worlds with exactly the same pasts making exactly the same effort and getting lucky in one world and not the other. Exact sameness (or difference) of possible worlds is not defined if the possible worlds contain indeterminate events of any kinds. And there would be no such thing as two agents having exactly the same *life histories* if their life histories contain indeterminate efforts or free choices. (171-2)

I find this statement puzzling. Unless I am grossly mistaken about how possible worlds operate, I can indeed imagine "the same agent in two possible worlds with the exact same pasts making exactly the same effort..." I believe I have just imagined it in the John1/John2 thought experiment.<sup>20</sup> It may be the case that I cannot explore the intricacies of John's brain and the quantum events that occur there, record them, and have them duplicated in the form of John2. That may forever be beyond us, but that is hardly required in this case. Whatever the results of an agent's mental processes, possible worlds can work in such a way that there

will always be another possible world that is identical with the agent's world such that even the indeterminate processes just so happen to have the same results. To simply state that this is not possible is not an adequate response to the objection.

I now move on to the second criticism of this sub-section. Indeterminism in Kane's theory serves the role of allowing for an agent to be ultimately responsible for her action. It is not clear why, however, Kane's placement of indeterminacy within the agent serves to make his theory exempt from traditional objections against indeterminism. Additionally, Kane would be hard pressed to show that indeterminism plays *any* role in explaining the freedom of the agent.

Consider a pair of magic dice. When rolled their outcome is indeterminate in the sense that even an omniscient being (who would be presumably informed of all the relevant facts and laws for the purpose of prediction) would not be able to predict which numbers will land facing up. If God does indeed play dice with the universe, these are the dice he would employ. The shopkeeper is faced with a dilemma – to overcharge or not. Which action the shopkeeper performs will be the result of an indeterminate effort of her will – an effort that can be represented by a toss of the magic dice. Kane has not shown that his theory gains anything by making the indeterminacy internal to the agent. Such placement of indeterminacy is an attempt to show that the agent has ultimate responsibility for her actions. However, what difference does it make if the magic dice belong to the agent or not? Regardless of to whom they belong the result of the throw will be equally indetermined and, because it is indetermined, not a result for which the agent can be held ultimately responsible. Whatever results would not be an action attributable to the agent, but rather a movement that just happened as a result of an indeterministic process. Whatever role

indeterminism plays in Kane's theory, it alone *does not* play the role of obtaining freedom for agents. For that, he must also look to folk psychological ascription and the phenomenological perspective.

The freedom described in Kane's theory is not gained solely via indeterminism and ultimate responsibility, but rather additionally through the claim folk psychology cannot be discarded. To discard folk psychology is to take something akin to the scientific perspective described by Kane. From this perspective, it is hard to attribute freedom to actions. Instead of actions performed by agents, there are only descriptions of movements of physical objects. It is only by virtue of describing certain movements in a certain way, namely, by describing the actions of agents from the phenomenological perspective, that freedom could possibly arise. Why did the shopkeeper charge fair prices for her products? It is the phenomenological perspective, that which employs folk psychological descriptions and ascription, which provides an acceptable answer: she did so because she had decided to be moral in the instance in question.<sup>21</sup>

Kane's dependence upon folk psychology is problematic because folk psychology *is compatible with determinism*. We can be determined in such a way that we attribute freedom to one another via folk psychological ascription. Kane's use of folk psychology strengthens the weakest part of compatibilist theories that have been criticized for being unable to account for how a determined motion can count as an action for which an agent can be held responsible. Using folk psychology does indeed make Kane's theory more plausible (from the point of view of the compatibilist, anyway) than those libertarian theories that are excluded by the free agency principle, but at the cost of bolstering

compatibilist intuitions and losing the ammunition that has traditionally been used against the compatibilist position.

Kane considers an objection in the eighth chapter of his book similar to what is made here (148). A hypothetical compatibilist complains that Kane is merely replacing one mystery – that of agent-causes, noumenal selves, or mind/body dualism – for another mystery – this time of indeterministic efforts of will described physically as indeterminate processes that are happening in the brain but phenomenologically as something that agents are doing. Kane agrees that this is so but notes that the second mystery is part of a larger problem of consciousness that, unlike the first mystery, is acceptable because Kane shares this mystery with compatibilists.

However, materialistic accounts of freedom (indeterministic or deterministic) that rely on folk psychology are in the same boat when it comes to the possibility of discovering the truth or falsity of determinism. Suppose it is discovered that determinism is false. Then compatibilist accounts can be slightly adjusted to take into account Kane's brand of indeterminacy. Alternately, suppose that determinism is somehow discovered to be true. Then incompatibilist accounts such as Kane's can be slightly adjusted.<sup>22</sup> The important point is that regardless of whether determinism is actually true or actually false, folk psychology can still be used because it is compatible with *either alternative*. Because of this, it cannot play the role of securing indeterministic freedom. Kane has espoused a position that is not primarily on the side of libertarians against compatibilists and hard determinists, but rather on the side of free materialists (whether compatibilists or incompatibilists) against libertarians and hard determinists.

A defender of Kane might respond that my objection has missed the point. There is a great difference between Kane and compatibilist accounts of freedom that rely upon folk psychology. This difference is the AP condition (or could have done otherwise). Whereas the compatibilist must give up a condition such as AP (taken in the libertarian sense, of course), Kane is able to employ it in such a way that allows for moral responsibility. This defender might say that quantum indeterminacy and folk psychology play vital but different roles in Kane's account. We have, via quantum indeterminacy, that things could have happened differently. What makes that *happening* into a *doing*, however, is folk psychology (and hence, we cannot be eliminative materialists). My response to this defender takes the form of a dilemma pertaining to what Kane means by folk psychology.

Consider two senses of folk psychology. The first sense is the simple idea expressed in the computational theory of mind that our beliefs and desires combine to determine actions. I refer to this as weak folk psychology (or WFP). The second sense is a much stronger sense that Kane seems to be getting at (I refer to it as SFP). The stronger version holds that there is an irreducible phenomenological component to our actions that we must take very seriously and serves to allow for moral responsibility. But does Kane appeal to WFP plus quantum indeterminacy or SFP plus quantum indeterminacy? Here the dilemma arises.

If he is appealing to WFP, then his account amounts to a version of the computational theory of mind attached to a random number generator (via quantum indeterminacy). This may give us a sense of being able to do otherwise, but it does not seem to be a sense that the libertarian would think is relevant to questions of moral responsibility.

If he is appealing to SFP, then I question whether quantum indeterminacy plays any necessary role in the equation. If SFP is enough to make an action a doing in a morally relevant sense for Kane, it should be enough to make an action a doing in a morally relevant sense for compatibilists.

## **2.4 Concluding Remarks**

In this chapter I have explained the indeterminist incompatibilist philosophy of Kane. In doing so, I have raised several objections specific to his account. Specifically, with his argument supporting the idea that the traditional sense of free will represents a significant freedom and with his attempt to provide an intelligible account of such freedom. In the next chapter, I discuss objections that pertain to indeterministic accounts of freedom in general. In doing so I discuss, when appropriate, how the objections relate specifically to the Kane's philosophy.

## **2.5 End Notes**

<sup>1</sup> Both Hobbes' and Bramhall's positions are outlined in Molesworth (1962).

<sup>2</sup> This is a challenge that Kane must answer, and he attempts to do so by trying to remove the confusion from the first horn of the dilemma and claiming that other libertarians are mistaken in trying to provide an answer to the second horn.

<sup>3</sup> Of course, we may also be held responsible for actions that we refrain from performing.

<sup>4</sup> The Martin Luther example is found in chapter six of Dennett (1984); the Black and Jones example is found in Frankfurt (1969), 835.

<sup>5</sup> Dworkin (1986), 424.

<sup>6</sup> van Inwagen (1983), 16, quoted in Kane (1998), 45.

<sup>7</sup> "Are We Free to Break the Laws?" in Lewis (1986), Vol 2: 291-8.

<sup>8</sup> Lewis, 297, quoted in Kane (1998), 49.

<sup>9</sup> *ibid.*

<sup>10</sup> This is so because we can imagine the laws of nature being different in such a way that they caused the individual to be radically different or caused her not to exist.

<sup>11</sup> The respective positions of Austin and Chisholm discussed here can be found in papers included in Berofsky (1966). They are Austin's "Ifs and Cans" (295-321) and Chisholm's "J. L. Austin's Philosophical Papers" (339-45).

<sup>12</sup> I argue in section 2.3.2 that the compatibilists can accommodate the data used by Kane in his dialectic of selfhood.

<sup>13</sup> I do not immediately see that the membrane metaphor succeeds. What materials the membrane of a cell allows in and out would be determined by the physical nature of the membrane so it does not seem to parallel an active choice made by an agent. However, it is just this type of worry, the worry that (say) her "membrane" is determined by forces outside of her control, that Kane argues leads the agent to the next part of the dialectic of self-hood.

<sup>14</sup> I argue in section 2.3.1 that this stopping point, placing free will as a higher state response to the dialectic of selfhood, is arbitrary.

<sup>15</sup> What Kane is minimally doing here is not allowing the libertarian to take the second horn of the libertarian dilemma. By doing so, does he take the first horn? Does he dissolve the dilemma?

<sup>16</sup> I first encountered the notion of a "just so story" in Daniel Dennett's *Elbow Room* (1984). I have since learned that Rudyard Kipling has written a series of "Just So Stories" for children that provide humorous answers to such questions as "How the Camel Got Its Hump" and "How the Leopard Got Its Spots." Additionally, in the biological sciences evolutionary explanations for behavior are criticized as being "just so stories." A most notable example of someone who lodges this type of criticism can be found in the work of the renowned zoologist Stephen Jay Gould.

<sup>17</sup> The problem of moral luck is discussed in Nagel (1979), 24-38.

<sup>18</sup> In reality this may never be so simple – there may always be other options available no matter how little the agent desires them. In this case, perhaps John as an additional minute desire to drop out of school and form a rock band. I have chosen to limit the options to two in order to make the example clearer.

<sup>19</sup> Waller (1988). Kane credits the following who have made a similar criticism: Thomas Talbot and Richard Double in correspondence; Mark Bernstein and David Blumfeld in discussion; and Galen Strawson in Strawson (1994), 19.

<sup>20</sup> The two Johns were identical until the first SFA mentioned in the example. They have had SFA's in the past, an equal number of them as a matter of fact, and each of them up until this point has had identical results. I can further imagine another individual who is identical to John1 and will be until they both die. What distinguishes the world of John1 from this other world could be an event that occurs in the future after both John's are dead. It is not necessary to offer an explanation for how John1 and John2 managed to be identical until the point at which they split (the first SFA in the example), it is enough to state that John1's history, whatever it may contain, determinate or indeterminate, can be cut and pasted, if you will, into another possible world.

<sup>21</sup> This is, of course, overly simplistic. The explanation would have to regress further – why did she decide to be moral in this instance? This latter question is problematic both for Kane and for compatibilists – for Kane because the introduction of indeterminism makes it more difficult to believe that the action resulted from a decision made by the agent; for the compatibilist because the explanation would eventually regress until a point of time before the agent was born, in which case how can he be held responsible? It is my contention that folk psychology plays a role in either case to attribute freedom.

<sup>22</sup> Kane's indeterminism could also function as a compatibilist position. For the compatibilist account, imagine the magic dice referred to earlier as only random with respect to human beings. The result of the throw would thus be determined and predictable (to God), but appear indeterminate to human beings.

### 3. Four Problems for Indeterministic Accounts of Freedom

In the previous chapter, I have discussed the libertarian philosophy of Robert Kane. While doing so, I touched upon the debate between compatibilist and indeterminist accounts of freedom. This chapter is divided into four main sections with each section corresponding to four objections to indeterminist accounts of freedom. In the first section, I discuss an objection taken from Galen Strawson that I have labeled “Strawson’s Challenge.” In the second section, I outline an objection raised by Thomas Nagel that he has dubbed the “Problem of Autonomy”. In the third main section, I discuss a strategy of Daniel Dennett’s that I refer to as the “Compatibilist Shift” with which compatibilists can develop theories of freedom and free agency that are as rich as those of the libertarians. In the final main section, I discuss Richard Double’s objection to libertarian accounts of agency, most notably those of Peter van Inwagen and Kane, which I refer to as the “Objection From Rational Explanation.” Additionally, I discuss how Kane is either susceptible to the objections or how he might respond to each of them where it is appropriate.

#### **3.1. The Strawson Challenge – No Place for Indeterminacy**

In “Libertarianism, Action, and Self-Determination”,<sup>1</sup> Galen Strawson considers whether libertarians can answer the skeptical objection that freedom is impossible regardless of the truth or falsity of determinism. If determinism is true, the objection holds, then our actions cannot be free because they are determined. If determinism is

false, our actions cannot be free because they result from a random process that defies complete explanation via the previous reason state of the agent. This is so because of the nature of self-determination that is necessary for freedom.

To illustrate, Strawson considers three versions of self-determinism. The first type of self-determinism considers an action as self-determined if it is a result of one's own choices, decisions, or deliberations. This statement of self-determination is compatible with determinism. One's deliberation and the outcome of the deliberation can be determined, but the deliberation must be performed by, and as such belong to, the agent who performs the action. Janice is offered a marijuana cigarette by a college roommate and must decide whether or not to accept it. Throughout her life she has met many people who smoke marijuana and has been disgusted by their apathetic attitude. However, she is genetically predisposed to be a risk taker and this results in her having a desire to experiment with the drug. She deliberates and her disgust is greater than her desire to take the risk. She decides not to accept the marijuana. In this example, Janice's action is the result of deliberation, but the result of the deliberation was determined by factors not necessarily under Janice's control. Though deliberation did take place, whichever desire was stronger is the desire that eventually won out. Because her disgust of drug users was greater than her desire to experiment, she could not but refuse the drug. Had the risk taking desire been stronger than her disgust, her deliberation would have ended by accepting the offered cigarette. Her action is a self-determining one, in this sense, because it is the result of her own deliberation even though the result of the deliberation is determined.

The second version, which Strawson attributes to the libertarian, is one of *true* self-determination. According to this version, “one is truly *self-determining*, in one’s actions, only if one is truly *self-determined*, and one is truly self-determined if and only if one has somehow *determined how one is in such a way that one is truly responsible for how one is*” (14). This type of self-determination can be demonstrated by slightly altering the example of Janice. In this case, Janice would be somehow responsible for her two opposing inclinations. By earlier actions for which she is responsible, she has become disgusted with marijuana users and developed an inclination to take risks. Again, whichever side is stronger will win out in the deliberation, but the action will be self-determining, in this sense, because Janice is somehow responsible for possessing the opposing desires and also responsible for their respective strengths.<sup>2</sup>

The third version of self-determination allows that one can be truly self-determining even if one is not responsible for how one is via some type of special intervention on the part of the agent. Regardless of whether or not Janice is responsible for her opposing desires and regardless of which desire is stronger, Janice is somehow able to intervene in the causal process resulting in an action that is not determined by previous events.

Both compatibilists and incompatibilists agree that the second type of self-determination is excluded by determinism. If I am caused to act because of circumstances antecedent to my birth, as determinism entails, then it cannot be true that I can be truly self-determined in this sense. At best, I can have self-determination only of the first type. This type of self-determination is rejected by the libertarian as not being self-determination at all. The skeptical question that remains to be answered is whether

true self-determination is possible if determinism is false. To be taken seriously as a libertarian, Strawson argues, it is not enough to simply renounce determinism and introduce indeterminism. The libertarian philosopher must also give an account of action production that locates the indeterminism in a non-trivial way that allows for free action. This is what I have dubbed “Strawson’s Challenge.” It is a challenge to libertarian philosophers to show where indeterminism could possibly be introduced such that it enables the agent to act freely.

### 3.1.1. Strawson’s Challenge

Strawson considers several possible responses to his challenge. Consider:

Assume that a particular action *A* performed by *a* is truly and fully explicable by reference to a reason-state *R* made up of desire(s) *D* and belief(s) *B* (or by reference to events characterizable in terms of desire and belief), while it also has an indeterministic input *X* among its antecedents. The question is, where can *X* be? (18).

The libertarian cannot respond by locating *X*, the indeterministic input, between *R*, the reason-state, and *A*, the action, because that would mean that *A* would not be truly and fully explicable by referencing *R*. *R* would determine that the agent order a salad, for example, but before the action can be expressed the indeterminate *X* causes the agent to order a burger instead. It is hard to imagine that an agent can be held responsible for such an action. The agent does not *do* anything, rather, something *happens* which interferes with what he intended to do.

Similarly, the libertarian cannot respond by claiming that *X* is unconnected with *R*. If so, then *A* would again not be truly and fully explicable by referencing *R*. Instead, *A* would be explained by referencing the determinate *R* and the indeterminate *X*. Strawson emphasizes that it is important for the indeterminacy not interrupt the connection between the *reason state* and the *action*. This is so because:

...it is specifically *qua* reasons-reflecting, reasons-determined things that actions must be shown to be free. If so, the indeterministic input allegedly necessary for free action cannot possibly be supposed to contribute to freedom either by interfering with or interrupting the determination of actions of reasons or because it is a contributory determining factor that is wholly independent of reasons for actions. So it can play a part only by playing a part in shaping or determining what the agent's reasons for actions are. (19)

A constraint upon libertarian accounts of free agency then is that they must be able to give an account of free actions such that the action is determined and explained by reasons possessed by the agent. Since *X* cannot occur between *R* and *A* and cannot occur separately from *R*, the libertarian must respond to Strawson's Challenge by claiming that *X* must be a factor in determining the *R* that determines the *A*. In this way it will still be true that *A* is determined by *R* and that the reason state *R* that determined the action *A* could have been otherwise. My reason state determined that I ordered salad, but because of *X* my reason state could have been different and, as such, I could have ordered hamburger. Placed in the exact same circumstance, I just may.

Given that the reason state *R* is made up of belief(s) *B* and desire(s) *D*, then *X* must play a role in determining either *B* or *D* or both. Strawson quickly discounts *X* playing a determining role in the beliefs of an agent. As rational creatures, we want our beliefs to be determined by and to accurately reflect truth and reality. When faced with the choice of whether to believe that I can leap off of a tall building and fly, I do not want the content of my belief to depend solely on me. I want my belief to be a true belief. When I order a salad I am, say, acting partly upon a belief that eating vegetables is good for me. I do not want that belief to be in any fashion arbitrary. Admittedly there may be cases in which it is advantageous and perhaps even desirable for our beliefs to not correspond to reality. For example, consider a member of a Nazi concentration camp who is able to completely delude himself into believing that the whole thing is an elaborate prank. Won't he laugh when the trick is revealed and his family and friends are

returned to him alive and well. In general, however, we do not object to the notion of our beliefs being determined. In fact, properly determined true beliefs are (with extreme exceptions, perhaps, but still) preferable to agents. Placing *X* within the realm of belief jeopardizes the relationship between our beliefs and reality and, as such, is not an acceptable locale for indeterminism.

Since *X* cannot be located within *B* (which precludes it from being located in both *B* and *D*), then *X* must be located within *D*. If *X* is to play a role at all, it must play a role in determining the desires of the agent. Thus far we have departed from talking of true self-determination in favor of mere indetermination represented by *X*. However, to show that *X* is not helpful when considered as a determining factor of *D* the notion of true self-determination must be brought back in. The question then becomes how the introduction of an indetermined cause of *D* could possibly help to establish true self-determination. The libertarian must be able to show how the agent is responsible for having those desires partly determined by *X*.

Thus far the type of determination dealt with has been mainly actions determined by reasons. Now consideration must be given to another type of determination, that of reasons being determined by agents. If I choose to order the salad because I desire to stick to my diet, I cannot be truly responsible for performing the action unless I was also somehow responsible for possessing the desire that determined the action. If the desire was the result of a choice on my part, that choice must have been made according to (and be determined by and explicable by) reasons for choosing the desire. If no reason can be given for possessing the desire, i.e., if the desire was the result of an indeterminate process, then I cannot be truly self-determining with respect to that desire nor the resulting action. If I did choose the desire for reasons, then the choice to accept the desire was determined by those reasons and I cannot be truly self-determined with respect to choosing the desire nor performing the action unless the reasons for choosing the desire were self-determined. The regress either extends deterministically beyond the time

I existed or comes to an arbitrary, undetermined, stopping point. Either way I am not truly self-determined.

Hence,  $X$  cannot play a causal role in the desires possessed by an agent. Since Strawson has shown earlier that  $X$  cannot occur in any other place, it appears that the libertarian is stuck. The argument against the inclusion of indeterminism is given twice by Strawson (16-17). The second, shorter, formulation is:

- 1) It is undeniable that one is the way one is as a result of one's heredity and experience.
- 2) One cannot somehow accede to true responsibility for oneself by trying to change the way one is as a result of heredity and 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 the attempt at change, will be determined by how one already is as a result of heredity and experience.

To be truly self-determined, an agent must somehow be able to choose her reason state without that choice being determined. Yet, as Strawson has shown, that choice cannot be simply indeterminate without violating true self-determination.

### **3.1.2. Robert Kane**

In the previous chapter I have discussed and criticized the libertarian philosophy of Robert Kane. It would be informative here to discuss how Kane seems to answer Strawson's Challenge. He does so by placing indeterminacy, the  $X$ , between the reason state,  $R$ , and the action,  $A$ . However, he attempts to do so without destroying the rationality of the action. He does so by rejecting what he refers to as one-way rationality in favor of plural rationality. In his philosophy, the agent can possess multiple competing reason states,  $R_1, R_2, R_3...R_n$ , each of which can result in an action for which the agent can be held responsible. The indeterminacy,  $X$ , determines which reason state will win out. In this way  $X$  has been introduced without threatening the rationality of the resulting

action and, since the reason state that produced the action was one for which the agent was responsible, the agent is responsible for the resulting action. Kane's placement of the *X* also serves as a challenge to the third step in Strawson's argument. The degree of success of one's effort to perform an action over another (choose one reason state over another) is precisely where *X* is introduced. The effort is an indeterminate effort.<sup>3</sup> Though this appears to satisfactorily answer Strawson's Challenge, Kane's account is troublesome for reasons I have already outlined in the previous chapter.

### **3.1.3. Special Intervention**

As mentioned earlier, there were three versions of self-determination. The first version was discarded because it was not consistent with libertarian ideas of freedom and, hence, was not *true* self-determination. The second version was examined and discarded because of an inability to place indeterminism within the decision making process of an agent. There remains the third version, however, in which an agent can be truly self-determining even if his reason state is not self-determined. This is so because of a special interventionary choice performed by the agent to do other than what the reason state has determined. This special choice is somehow not determined by the beliefs and desires which have (admittedly) been determined by forces outside of the agent's control. There I sit in the restaurant. My reason state determines that I order a salad but, before the order to the waiter can escape my lips, I suddenly intervene in my action determining process to order hamburger. But why did I do this? Did another waiter pass by to deliver a hamburger to another customer and the smell of it made my mouth water? Did I remember that there had recently been trouble with illegal and dangerous pesticides being used on crops of lettuce? Can I explain why I altered my choice?

If the answer is yes, then it seems that I made the special interventionary choice based upon *another* reason set of beliefs and desires and, as such, it was equally as determined as my original intention of ordering the salad. If no explanation is possible for the altering of my choice, then we are faced with explaining how the decision is one for which I am responsible as opposed to one that just happened.

Strawson additionally makes the stronger claim that if this were somehow true, if the special intervention existed and somehow resulted in a choice for which we are responsible, then we are not free with respect to any action for which we can provide a full and rational explanation. If we are free only in virtue of the special intervention, then it seems we are only free when we exercise this special power. Though I agree with his general argument against this type of indeterminism, I do not think he is justified in making this stronger claim. The libertarian who holds this position can argue (again, provided that it is agreed that the special intervention does result in a free choice for which the agent is responsible) that in cases where no intervention is made the power to intervene was present but not exercised. Or rather, it was exercised but its result did not conflict with what was determined by the reason state. My environment has conditioned me to strongly desire hamburgers. At the restaurant, when faced with the choice of hamburger or salad, my reason state determines that I choose salad. Additionally, I exercise my special interventionary powers. I can either choose to go against my reason state and order the salad or choose to order the hamburger in agreement with my reason state. In either case, I am responsible (*ex hypothesi*) for the resulting action. Still, Strawson has done enough to show that libertarian positions that rest upon indeterminism have much to explain.

### **3.1.4. Leibnizian Free Will**

Finally, Strawson considers a Leibnizian view on libertarian free will in which the reason state plays an influential role in the determination of action without necessarily being sufficient for causing the action. These reasons can affect one's decision making process without wholly determining the result of the process. The argument against this view is along the same line as Strawson has expressed earlier. If the agent is able to do other than what his reason state dictates, he must do so in virtue of further beliefs and desires he possesses. If this is so, however, he is not truly self-determining and not truly responsible for the resulting action. If such further beliefs and desires are absent, then the resulting action is, rationally speaking, random.

### **3.1.5. Conclusion**

Though the type of freedom for which true self-determination is necessary seems obviously impossible, Strawson argues that it is important to examine because it is precisely the type of freedom that most people commit themselves to in everyday talk. As such, they have a "crucial role in structuring our attitude to the notion of freedom" (28). However, libertarian philosophers must provide an acceptable answer to Strawson's Challenge to show that indeterminism can aid in actions which are truly free.

## **3.2. Nagel's Problem of Autonomy**

In "The Problem of Autonomy",<sup>4</sup> Thomas Nagel examines the problem of free will. The problem expresses itself when we take an objective view of ourselves and others. In doing so, we seek to explain our actions and the actions of others causally. My not wanting to get caught caused me to not cheat on the exam while Mary's wanting to

pass the exam at any cost caused her to risk cheating. Initially, Nagel argues, considering actions from the objective perspective seems to increase our freedom. We do not simply act, but rather deliberate and consider various courses of action before deciding to act. Or, when simply acting, we do so based upon a character that we have previously chosen. If taken further enough, however, the objective view destroys what it initially enhances. When we step far enough outside of ourselves it is hard to see ourselves and others as agents rather than as parts of nature. My wanting not to get caught results from having suffered severe consequences for cheating as a child. Mary's wanting to pass at all costs is a result of a demanding parent who finds anything less than an A by Mary to be unacceptable.

Nagel considers two aspects of the problem of free will. The first is the problem of autonomy and the second is the problem of responsibility. Traditionally, the problem of free will has been discussed relative to the problem of responsibility. If Mary's decision to cheat was caused by conditions over which she could exercise no control then we cannot hold her responsible for the action. The action was not performed freely. This undermines the reactive attitudes that are conditional to the attribution of responsibility. It is of no use to resent Mary for cheating and getting an A because it was not an action for which she can be held responsible. This is true even though we may not be able to help having the resentful feelings.<sup>5</sup>

Although the second problem is generally considered as *the* problem of free will, Nagel argues that the problem of autonomy is equally threatening to our conception of free will. The problem of autonomy is the fear that the idea that our own actions are

freely performed by us as agents is merely an illusion. We really do not *act* at all, but rather what we do is only what happens through natural and physical law.

The problem of autonomy results in the hopeless situation of wanting something impossible. This is a result of two feelings. On the one hand the feeling of unease when we try to take the objective perspective to heart. Though the external view cuts away the support for our autonomous feelings, “the unstrung attitudes don’t disappear...despite their loss of support” (35). No amount of stepping outside of oneself will cause these feelings about our autonomy to cease. These feelings, however, are certainly not proof that we are free. To be such a proof, it needs to be shown why they can serve as explanations of our freedom rather than simple subjective impressions of how action seems to the agent (39). On the other hand, if we fail to consider the objective perspective then we cannot allay the feeling that, if we were to look from a distant enough perspective, an agent’s actions are helpless and not something for which he can be held responsible (35). Nagel argues that no attempt at eliminating the objective perspective can alleviate this fear.

The problem lies when we try to give a coherent account of what these internal impressions of autonomy amount to. Nagel argues that no attempt to provide such a coherent account of the internal view of action has been provided. This is especially troublesome because it is just this view that is in danger of being discounted by the objective perspective. “When we try to explain what we believe which seems to be undermined by a conception of actions as events in the world - determined or not - we end up with something that is either incomprehensible or clearly inadequate” (35). From the inside it seems that we have alternate possibilities open to us. I can choose to go to

class or get some extra sleep. Whichever possibility is chosen is actualized by my choice. When considered externally, however, it seems that only one of the actions was actually possible, the one that actually occurred. My eventual decision to go to class and forgo the extra sleep is explainable as, say, an expression of my character. A character that I could not have freely chosen.

Though we accept a subordination of subjective appearance to objective reality in other areas, we cannot accept it in the area of free will. Nagel argues that this is so because action is too ambitious - the idea of our autonomy is not simply a feeling but rather a belief. We cannot regard our feeling of freedom as being mere appearance without giving up this belief. Though he considers this belief to be unintelligible, Nagel offers a description of what our ordinary conception of autonomy is. It is the belief that antecedent circumstances, including the character of the agent, are not sufficient to determine all action. Somehow, the agent can choose to break the causal chain by making a choice that is both inexplicable by antecedent causes yet remains a choice for which the agent can be held responsible. The final explanation of the resulting action is not causal but rather intentional. This intentional explanation is comprehensible only from my point of view from which “[m]y reason for doing it is the *whole* reason why it happened, and no further explanation is either necessary or possible” (37).<sup>6</sup>

The external view does not allow for intentional explanations, but rather only causal explanations. The absence of causal explanation then amounts to having no explanation at all for why an action occurred. For the libertarian to defend his notion of freedom, he must require that such intentional explanation be acknowledged. The problem is that this only gives a correct surface description of our “prereflective sense of

our own autonomy” (38). Intentional explanation collapses when examined closely, however, because it can be given for any resulting action. Consider Sally, a high school senior having to choose between two prospective dates to the prom. The first choice, Biff, is a wonderful physical specimen but lacking in brains and charm. Peter, the other possibility, is smart and considerate but not much to look at. Sally undergoes an internal struggle - should she choose the smart guy or the attractive guy?

A causal explanation can be given from the external perspective regardless of whom she chooses (since, in reality, there was no choice). If she chooses Peter it is because, say, her older sister married a nice guy like him and is deliriously happy. If she chooses Biff it is because, say, she has been influenced by the plight of her mother who did not have any fun in her life before settling down. The action does not determine the causal explanation, but rather we can infer which causal explanation has determined the action.

From the intentional perspective, the causes mentioned previously contribute to the resulting action without determining it.<sup>7</sup> Space is left for Sally to choose the reasons for which to act. Her possession of an appropriate set of reasons,  $R_1$ , for choosing Biff and another appropriate set of reasons,  $R_2$ , for choosing Peter render whichever choice she makes intelligible, but from the internal perspective they cannot explain why she found one set of reasons more appropriate than the other. To do so intelligibly she would need to appeal to *other* sets of reasons – an appropriate set of reasons,  $R_3$ , for choosing to accept  $R_1$  over  $R_2$  and an appropriate set of reasons,  $R_4$ , for choosing  $R_2$  over  $R_1$ . Additionally, the choice of either of those reason sets would have to be made based upon a further set of reason states. Either the regress is infinite or there is an arbitrary stopping

point. Hence, intentional explanation collapses because it cannot explain why the agent chose one set of reason over other, equally intelligible sets of reasons.

### **3.2.1. Robert Kane**

Nagel's problem of autonomy serves as a serious objection to libertarian philosophies in general and Kane's philosophy in particular. Kane attempts to answer this type of problem by arguing that we cannot be an eliminative materialist when it comes to human action. In Nagel's terminology, we cannot give up the internal perspective for the external perspective. For Kane, both the internal perspective and the external perspective are equally valid as explanations for our actions. These intentional explanations<sup>8</sup> cannot be given up without writing free will out of the picture. Kane rightly points out that this is a problem that all materialistic accounts of free agency share, both compatibilist and libertarian. This is good enough for Kane's purposes, but not a sufficient response to the problem of autonomy. Nagel's problem of autonomy asks the libertarian to explain *how* an agent can hold onto his internal feeling that his action is free when faced with the problems that arise from taking an objective view of his action. Kane does not do this – there is no *how*, merely a *somehow*. This problem will be solved, according to Kane, when the greater problem of consciousness is solved.

### **3.3. Dennett's Compatibilist Shift**

In "On Giving Libertarians What They Say They Want,"<sup>9</sup> Daniel Dennett engages in a bit of subterfuge by offering the libertarian philosopher a framework in which a motion that results from indeterminacy can correctly be viewed as an *action*. I'll get to the subterfuge later, but first I'll describe how he proceeds.

Dennett imagines placing a person in an “answer box”. This box has two buttons (a “yes” button and a “no” button) and two pedals (again, a “yes” and a “no”). Also within the box is a display screen that says either “use the buttons” or “use the pedals”. The subject is then asked a series of ten simple yes or no questions and responds by either using the buttons or the pedals depending upon what the display instructs. However, whether the display says “use the buttons” or “use the pedals” is based on an indeterminate process (perhaps, via a radium randomizer).

Dennett then considers whether a physicist could in principle predict the subject’s behavior. The physicist is given foreknowledge of the initial conditions of the subject as well as the answers to the ten easy questions. Because of the introduction of indeterminacy, the physicist could at best answer with a series of “if...then” statements. If the display says “use the pedals” when question one is asked, then the subject will press the pedal which corresponds to “yes”. (This is very general. The physicist’s actual prediction would involve the motion of atoms of some type causing motion of other atoms which result in macro movement. The more general description demonstrates the point well enough). Dennett compares the results of the physicist with the results of an intentionalist who tries to perform the same task. The intentionalist can, upon reading the questions, predict that the subject will answer “yes” to questions (say) 1, 2, 7 and 9 and will answer “no” to all others. It is important to note that there are no “if’s” or “maybe’s” in the prediction of the intentionalist. This serves as an instance in which indeterminacy is placed within a system without entailing that accurate predictions can not be made about the performance of the system. The predictive power of the physicist and the intentionalist are equivalent. But in our everyday talk about predicting, so Dennett

argues, we are not interested in the type of predictions that occur on the purely physical description provided by the physicist. Rather, we are interested in being able to predict actions. This type of prediction can be made from the intentional stance. Dennett then places the “answer box” within an agent in an attempt to discern what would result.

Suppose that I decide that I want to insult my neighbor and that the decision was a result of a determinate process. The act of insulting my neighbor, however, can be accomplished in an infinite number of ways. For the sake of the example, I favor no particular insult over any other and any choice of expression is arbitrary. Dennett considers the effect of placing the “answer box” at this point. I choose to perform an action, but how the action is actually expressed is (insofar as I don’t care about how it is done) decided by an indeterminate decision making process. The “answer box” functions merely as tie-breaker. In this way indeterminism can be introduced yet accurate predictions can be made by the intentionalist as to human behavior. Dennett argues that this does not give the libertarian what he says he wants, however. “The libertarian would not be relieved to learn that although his decision to murder his neighbor was quite determined, the style and trajectory of the death blow was not” (49).

Having demonstrated that indeterminism could function within the agent without interfering with the intentionalist's predictive power, Dennett turns to the task of placing the indeterminism in a place that would be acceptable to the libertarian. To do so, he discusses the role time pressure has in the decision making process of humans. It may be that if we were not under time constraints then we would always act rationally (and determinately). Consider the case of a high school senior deciding which college to attend. If there were no time constraints, it would be arguably possible for her to take

into consideration each and every possible advantage and disadvantage that each college offers. Time is a factor, however, and she cannot come close to exhausting these considerations. The student must rely upon a heuristic decision making process. Dennett argues that whichever (of the exhaustive) considerations the student does base her decision upon are brought about through an indeterminate process. The agent makes her choice (a choice for which she is responsible) based upon reasons that are in part generated through this indeterminate process. After choosing one school over the other, she can slap herself in the head for failing to think of a relevant consideration under which (had it occurred to her previously) she would have made a different choice. She chose school A, for example, but did not take into account that her uncle teaches at school B and could help show her the ropes. Had she remembered her uncle at the time of choice, she would have chosen B. Nevertheless, it was still the agent's choice. Dennett argues that, to the extent that indeterminism can make sense in theories of agency, the indetermination must occur in a place such as this. Otherwise indeterminism would be installed in a "harmless place by installing it in an irrelevant place" (49).

### **3.3.1. Absolute and Relative Randomness<sup>10</sup>**

Earlier I mentioned that Dennett would engage in subterfuge when he describes how indeterminacy can be introduced into the decision making process of an agent while still having the resulting choice and action being one for which the agent can be held responsible. The subterfuge rests upon an ambiguity concerning the type of randomness entailed by indetermination.

In the sense of indetermination that Dennett has used in his "answer box" example, the resulting expression of the action was what I shall call *absolutely random*.

In the answer box example, the absolute randomness was generated via a radium randomizer. From the scientific perspective, which way I insulted my neighbor was not predictable in principle because of the nature of the indeterminacy introduced. This is because the answer box was connected to a “radium randomizer” that, presumably, is not predictable in principle. However, *absolute randomness* in the sense I am employing is a stronger notion than this. Not only is it not predictable in principle for humanity, but not even an omniscient being could predict what will result. This is because the indeterminacy is ontological, as opposed to epistemic, in nature.

Consider a second form of indeterminacy, that which is *relatively random*. This is the weaker and more natural form of unpredictability in principle. Consider if Einstein was right in claiming that God does not play dice with the universe. In this case there could be no *absolute random* actions, i.e., actions that even an omniscient observer could not predict. Because of the nature of the universe it could be the case that certain events are destined to remain outside of the explanatory power of humanity. These events, though epistemically unpredictable in principle for humanity, are nonetheless ontologically determined and predictable by the omniscient observer. Lacking omniscience, some events that are not absolutely random will always *appear* absolutely random, that is, epistemically random relative to us. These events are not predictable in principle with respect to beings with finite knowledge but completely predictable by an omniscient being.

### **3.3.2. The Compatibilist Shift**

Of the two types of randomness I have just discussed, libertarian philosophers would appeal to absolute randomness to explain freedom of the agent. This is so,

obviously, because relative randomness is compatible with determinism and libertarian freedom is not. Dennett's point can be seen, however, when what I refer to as a "Compatibilist Shift" is performed on indeterministic libertarian theories of agency. The shift occurs by first locating the role that absolute randomness plays in the indeterministic libertarian theory of free agency. Once located, *absolute* randomness is replaced with *relative* randomness. Theories of agency or freedom that are incompatibilist thus can *shift* and become theories of agency or freedom that are compatibilist.

The Compatibilist Shift gives rise to a dilemma for libertarians who rely upon absolute randomness to gain freedom. On the one hand, the libertarian could admit that the "shifted" compatibilist theories do provide an adequate account of freedom and free agency. However, the libertarian cannot admit this and still be a libertarian. Alternately, the libertarian could die in the ditch for absolute randomness and argue that no account of freedom or free agency that uses relative randomness without absolute randomness can truly capture libertarian freedom. This seems odd because the "shifted" compatibilist accounts are just as rich and as lively as their libertarian counterparts excepting the small change that occurs. If the libertarian was able to produce a coherent theory of agency that uses absolute randomness (and this is a large supposition), then the compatibilist can adopt it and substitute absolute randomness with relative randomness. At the very least, the libertarian should admit that such shifted accounts are more acceptable than other compatibilist accounts that do not employ relative randomness.

To the extent that we cannot prove that the randomness in the world is absolute or merely relative, the theories of agency (indeterminate libertarian ones and their "shifted"

counterparts) would function in an exactly similar manner. Dennett discusses the chance of such a discover in a later work, *Elbow Room*:

...it is extremely unlikely, given the complexity of the brain at even the molecular level, that we could ever develop good evidence that any particular act was...a large-scale effect of a critical subatomic indeterminacy. So if someone's responsibility for an act did hinge on whether, at the moment of decision, that decision was (already) determined by a prior state of the world, then barring a triumphant return of universal determinism in microphysics, the odds are very heavy that we will never have *any* reason to believe of any particular act that it was or was not responsible. The critical difference would be utterly inscrutable from every macroscopic vantage point, and practically inscrutable from the most sophisticated microphysical vantage point imaginable. Some philosophers might take comfort in this conclusion, but I would guess that *only* a philosopher could take comfort in it.<sup>11</sup>

In this passage Dennett doubts that we will ever learn which type of randomness actually holds, whether it be what I have labeled absolute randomness or what I have labeled relative randomness. Quantum physics, for example, teaches that there are indeed undetermined processes on the sub-atomic level. This does not preclude us from learning in the future (through advances in technology, say) that the indeterminacy is merely relative and has a more foundational, determined, explanation. Wherever the investigation into the nature of the universe ends, there is always the possibility that there is another more foundational layer, either undiscovered or undiscoverable, that may not correspond to the previous scientific theories.

Libertarian philosophers can agree with this while not agreeing that the problem of free will is epistemic in nature. They can agree that we will never know what the truth actually is while still arguing that the libertarian theories come closer to capturing our everyday notions of freedom and its link with responsibility. However, at the very least

Dennett has shown the libertarian philosopher that a compatibilist theory of agency may not be as counter-intuitive as once thought.

### **3.3.3. Robert Kane**

The Compatibilist Shift is a useful tool for compatibilists who wish to criticize libertarian accounts of freedom and free agency like Kane's that rest upon absolute randomness.<sup>12</sup> I have used the strategy of the Compatibilist Shift to show that Kane's account serves to strengthen the intuitions that underlie compatibilist accounts of freedom and free agency in the previous chapter. This was done when I noted that a compatibilist version of Kane's account could easily be given by substituting a type of relative randomness for the absolute determinism of the "magic dice."<sup>13</sup> I do not further develop it here. What I do note here is that Kane claims that the answer to the problem of autonomy is found in the answer to the problem of consciousness. It would seem odd, then, if Kane stuck to his libertarian guns were the problem of consciousness to be solved in a way that favors compatibilism. Most compatibilists would not be in a similar position were the problem of consciousness to be solved in a way that favors incompatibilism. This is so because most compatibilists argue that freedom of the will is compatible with both determinism and indeterminism.<sup>14</sup>

### **3.4. Objection from Rational Explanation**

In "Libertarianism and Rationality,"<sup>15</sup> Richard Double criticizes defenses of libertarianism provided by Peter van Inwagen and Kane.<sup>16</sup> In doing so, Double argues that neither account leaves room for the reasonableness of libertarian free choices. I refer Double's argument as the "Objection from Rational Explanation."

Both van Inwagen and Kane are concerned that compatibilist accounts of freedom are not sufficient for attributing freedom and responsibility to agents. Van Inwagen's popular argument against compatibilism is the worry that what we do is not "up to us" because what we do is determined by the laws of nature and events of the remote past. Kane expresses similar misgivings by appealing to a notion of "ultimate responsibility" that is lacking in compatibilist accounts of freedom.

Both libertarians appeal to indeterminism to escape their worries but differ in where the indeterminacy is placed. Kane locates the indeterminacy in the psychological states that occur *before* a choice is made. Philosophers who locate indeterminacy before the moment of choice are referred to by Double as "Valerian libertarians."<sup>17</sup> Non-Valerians, as typified by van Inwagen (as well as Kant, Taylor, and Chisholm), hope to achieve freedom by placing the indeterminacy *at the moment* of choice while keeping all previous psychological factors the same. Both accounts are discussed in light of Kane's Condition of Ultimate Dominion (CUD) (59).<sup>18</sup>

CUD has two requirements for a free choice. The first is that "the agent's making the choice rather than doing otherwise...can be explained by saying that the agent rationally willed at *t* to do so". The second, that "no further explanation can be given for the agent's choosing rather than doing otherwise...that is an explanation of conditions whose existence cannot be explained by the agent's choosing or rationally willing something at *t*" (ibid.). Double refers to the first requirement as the requirement for rational explanation and the second as the requirement for indeterminacy.

Double takes CUD to be an accurate description of the type of requirement that must be met by any satisfactory libertarian view. He argues first that van Inwagen's

account of free will is a clear violation of CUD and second that Kane's account, while not violating CUD, does so at the price of producing "a theory that is weak on its rationality commitment without enjoying the incompatibilist advantages of van Inwagen's view" (ibid.).

### 3.4.1. Peter van Inwagen

Van Inwagen's account of free will is illustrated by his example of a thief who sometimes while stealing money remembers the face of his mother as she lay dying. On her deathbed he had promised her to give up his thievery and lead an honest life. Though the memory of his mother appears every time the thief steals, it is only successful in preventing him from stealing roughly half of the time. In each case, however, it is true that given the same psychological state and antecedent circumstances there exist some possible worlds in which the thief proceeded with the theft and others in which he did not.

Double argues that van Inwagen's account and other non-Valerian accounts do not satisfy CUD's rationality requirement. To show why, Double introduces "the Principle of Rational Explanation (PRE)" (60) which states that:

Citing a person's reasoning process *R* rationally explains a choice *C* only if the probability of *C* given *R* is greater than the probability of not *C* given *R*.

PRE is seen as a *minimal* requirement for rational explanation. Given my belief that the pursuit of higher education is noble and my strong desire to be noble, my choice to pursue a higher education can be rationally explained. However, this belief and desire would not serve as a rational explanation for my quitting school and joining the circus. In the first case, the belief and desire have, say, a higher than .5 probability of

determining my choice. In the second case, the belief and desire are irrelevant to the choice so do not serve as a rational explanation.

Returning to van Inwagen's thief, Double considers whether PRE is met. It is given that the thief's psychological state is the same regardless of whether he continues with his theft or turns away empty handed. Consider this psychological state  $P$ . However, if  $P$  is to serve as a rational explanation for the resulting action, it must be that  $P$  more likely determines one particular action over another. Given that  $P$  is the same in every case, there are three possibilities. Either  $P$  tends to determine that the thief continues with his theft, or  $P$  tends to determine that the thief discontinues his theft, or  $P$  equally tends to determine either outcome. In the first two cases,  $P$  can only serve as a rational explanation (via PRE) in the event that the action that has a higher probability to result actually results. If the less probabilistic action occurs then  $P$  would be precluded as a rational explanation by PRE. In the latter case, in which  $P$  equally tends to determine each outcome,  $P$  cannot serve as a rational explanation for either resulting action. Since  $P$  is neutral with respect to which action to undertake,  $P$  is irrelevant to the rational explanation of that action. It may be the case that rather than equally tending to determine each outcome  $P$  instead sometimes favors one (when the recollection of his mother's face is particularly vivid) over another (when the thief is starving). In this case, however, it is hard to argue that the thief is in the same  $P$  when committing either act. Rather two different psychological states would be involved,  $P_1$  and  $P_2$ .  $P_1$  would serve as a rational explanation for action  $A_1$  (why the thief did not steal) and  $P_2$  would serve as a rational explanation for action  $A_2$  (why he did steal). But this is certainly not what the non-Valerian is after.

### 3.4.2. Robert Kane

Though damning for the non-Valerian, Kane's account of free practical choices is in accord with both PRE and CUD. Consider again psychological states  $P_1$  and  $P_2$  of van Inwagen's thief. In Kane's account,  $P_1$  would serve as a rational explanation of  $A_1$  and  $P_2$  would serve as a rational explanation of  $A_2$ , but whether  $P_1$  or  $P_2$  is actually the psychological state of the thief is undetermined.<sup>19</sup> Therefore it is true that his psychological state can serve as a rational explanation for his action under PRE (and the first requirement of CUD) and that there would be no further explanation given that falls outside of the agent's choice or rational willing (or the second requirement of CUD). This is so because whether the face of the thief's mother appears prominently or not is undetermined. However, if her face does so appear the agent can do no other but refrain from the theft. Should her face not so appear then the thief can do no other but continue the theft. In either case, however, it is true that the thief could have done otherwise because his psychological state could have been otherwise.

Double offers three strong criticisms for Kane's view. The first is an objection raised by Dennett concerning Valerian accounts of free will.<sup>20</sup> It is hard to see how the agent can be held truly responsible considering that the agent is *not* responsible for the psychological state that determines the action. Once the undetermined event occurs then the action flows in a determinate fashion. Because of this indetermination does not help the rationality of the choice. How can the thief be responsible for the theft if his doing so is determined by the lack of a psychological state of which he has no control?

Secondly, the indeterminate events make no contributions to the psychological state that determines the action unless the agent first examines and interprets said events.

In the case of the thief, it is true that the strong image of his mother's face occurs indeterministically. However, once it does occur it is up to the thief to decide whether this event plays a determining role in his decision making process. If this is so, however, then the decision whether or not to let the event affect him must also be a choice for which a rational explanation is given. This second rational explanation can only be given via an undetermined event that the agent must also decide to accept or reject. Hence, an infinite regress of rational explanations and undetermined events.

The final criticism is that the Valerian approach does not admit categorical contra-causal freedom in the way that non-Valerian approaches do. Given the circumstances an agent was in at the moment of choice, it is simply false that the agent could have chosen other than what he does in the categorical contra-causal sense. Rather, he can only have chosen otherwise provided that his psychological state had resulted from a different undetermined event. This contra-causal freedom is no more than the type espoused by compatibilist accounts of freedom.

### **3.5. Concluding Remarks**

In this chapter I have discussed several problems that indeterminist accounts of freedom and free agency face. I am not aware of any account that can successfully fend off each of these objections. However, in the next chapter I pretend that one such account exists in order to discuss some further problems that it may face.

### **3.6 End Notes**

<sup>1</sup> O'Connor (1996), 13-32. All page numbers in this chapter refer to this O'Connor text.

<sup>2</sup> This example is similar to Dennett's example of Martin Luther (at least as Kane employs it) that I have discussed in the previous chapter.

<sup>3</sup> This is a simplified account of his philosophy, for a more fleshed out version please see chapter two.

<sup>4</sup> O'Connor (1996), 33-42.

<sup>5</sup> See Peter Strawson's "Freedom and Resentment" in Watson (1982).

<sup>6</sup> Roderick Chisholm argues similarly in "Human Freedom and the Self" when he claims that "in one very strict sense of the terms, there can be no science of man." Watson (1982), 24-35.

<sup>7</sup> This is very similar to the Leibnizian account discussed in the previous section of this chapter.

<sup>8</sup> Or, as he calls them, phenomenological and folk psychological explanations.

<sup>9</sup> O'Connor (1996), 43-56.

<sup>10</sup> Absolute randomness and relative randomness are my terms, not Dennett's.

<sup>11</sup> Dennett (1984), 136.

<sup>12</sup> As opposed to those that rely upon mysterious forms of causation or entities which Kane has outlawed via his Free Agency Principle.

<sup>13</sup> The magic dice are discussed in section 2.3.2 of this work.

<sup>14</sup> At the very least, they argue that freedom does not imply indeterminism.

<sup>15</sup> O'Connor (1996), 57-65

<sup>16</sup> Double refers to Kane's *Free Will and Values*, an earlier work than what I have discussed in chapter two. Van Inwagen's position is taken from *An Essay on Free Will*.

<sup>17</sup> After a citing by Daniel Dennett of Paul Valery's claim that invention is selection among choices that occur to one randomly.

<sup>18</sup> Kane's CUD is an earlier precursor to condition UR that is discussed in the previous chapter.

<sup>19</sup> This is so in Kane's later work because the effort to place one *P* over another is an indeterminate effort.

<sup>20</sup> In Dennett (1978), 297-98.

## 4. The Limits of Indeterministic Freedom

Even if libertarian theories of freedom and agency like Kane's could be made intelligible, they face the problem of not allowing us to be as free as we may like. Even if we possess libertarian freedom, it may also be true that in almost every instance we are not able to do other than what we do. This again raises the problem of moral responsibility because we intuitively do not think that we should hold each other responsible if we could not do other than what we in fact do. In this chapter, I examine three articles, two by Peter van Inwagen and one by John Martin Fischer and Mark Ravizza. Van Inwagen argues that even if libertarians are correct, we are not able to do otherwise as often as we think but are rather only able to do so in special circumstances. Fischer and Ravizza object and argue that circumstances in which we are able to do otherwise occur more often than van Inwagen claims and also argue that van Inwagen's view has unacceptable consequences for moral responsibility.

### **4.1. Introduction**

Peter van Inwagen begins "When is the Will Free?"<sup>1</sup> by stating that his argument will depend upon thinking of the problem of free will and determinism in this "classical tradition" in which an agent is ascribed "free will" if at some time it is true that he can choose to pursue either of incommensurable alternatives available to him. Within this tradition, van Inwagen argues that philosophers come to be libertarians by (implicitly or explicitly) relying upon a rule of inference that is something like what he refers to as "Beta."

A central problem in the debate over the problem of free will lies in the interpretation of the phrase “could have done otherwise” or “can do otherwise”. For van Inwagen, there is a single interpretation which compatibilists and incompatibilists alike agree upon. Under this interpretation, an agent truly “can do otherwise” or “could have done otherwise” only if his doing so does not either cause a previous state of the universe to be altered or cause a law of nature to be broken. In other words, all past events being equally unalterable and under the same laws of nature, the agent can perform either of two incommensurable acts. A central thesis of his paper is that though compatibilists can argue that there are numerous occasions upon which agents are able to do otherwise, incompatibilists must conclude that being able to do otherwise is a relatively rare condition.

In “When the Will is Free”,<sup>2</sup> John Martin Fischer and Mark Ravizza criticize what they refer to as the “restrictive incompatibilism” or “restrictivism” of Peter van Inwagen. Fischer and Ravizza describe van Inwagen’s position in this manner because if van Inwagen’s position is true (and determinism false) then incompatibilists must accept radical restrictions on one’s ability to do otherwise. Fischer’s and Ravizza’s two main criticisms are first that one can be an incompatibilist without being committed to the restrictivist position, and second that restrictive incompatibilists cannot provide a satisfying theory of moral accountability while still remaining within the classical tradition.

In “When the Will is Not Free”,<sup>3</sup> Peter van Inwagen responds to three criticisms leveled by Fischer and Ravizza. The first is that incompatibilist do not have to accept the validity of Beta; the second that the validity of Beta does not entail that we are able to act

otherwise as seldom as van Inwagen claims; and the third that restrictivism entails that we can seldom if ever be held morally accountable for what we have done.

This chapter is broken into three main parts to correspond to three main bones of contention between van Inwagen on one side and Fischer and Ravizza on the other. Additionally, each main section is broken into three sub-sections. In the first sub-sections, I outline van Inwagen's position as described in "When is the Will Free?" (this article will be referred to as VI(a) for van Inwagen (a)). In the second sub-sections, I discuss the objections raised by Fischer and Ravizza in "When the Will is Free". In each third sub-section I describe van Inwagen's response to Fischer and Ravizza as described in "When the Will is Not Free" (this article will be referred to simply as VI(b) for van Inwagen (b)) while both criticizing said responses and also providing independent arguments to show that Fischer's and Ravizza's objections do not hold.

The three bones of contention are theses proposed by van Inwagen in VI(a). They are as follows:

- (1) In order to be an incompatibilist one must accept as valid a rule of inference that van Inwagen has labeled "Beta";
- (2) Beta implies that we are seldom if ever able to otherwise; and
- (3) The previous claim does not entail that agents can only seldom if ever be held morally accountable for their actions.

#### **4.2. Incompatibilism and Beta-Like Rules**

Van Inwagen argues that incompatibilists must accept Beta as a valid rule of inference. In this section, I discuss van Inwagen's argument as well as Fischer's and Ravizza's objection. They attempt to provide an argument for incompatibilism that does

not rely upon Beta but rather employs two fixity principles (fixity of the past and fixity of the laws of nature). Van Inwagen and I both argue that their attempt to provide an argument for incompatibilism using the fixity principles fails because we cannot derive the absurd conclusion that we are unable to do other than we do from the two fixity principles.

#### 4.2.1. van Inwagen – Rule Beta

Van Inwagen claims that, generally, persons become incompatibilists because they are convinced by an argument which relies on two rules of deduction involving  $p$  (which stands for any true proposition) and the operator ‘N’ where ‘ $Np$ ’ stands for “ $p$  and no one has, or ever had, any choice about whether  $p$ ” (224).<sup>4</sup> These two rules are:

Rule Alpha: From  $\Box p$  deduce  $Np$ . (Where ‘ $\Box$ ’ represents “standard necessity”: truth in all possible circumstances.)

Rule Beta: From  $Np$  and  $N(p \Rightarrow q)$  deduce  $Nq$ . [Where “ $\Rightarrow$ ” represents a conditional.]

Let ‘ $P$ ’ represent any true proposition, ‘ $L$ ’ represent the conjunction into a single proposition all of the laws of nature, and ‘ $P_0$ ’ represent a proposition that gives a complete and accurate description of the whole world at an instant in the past before human life had evolved. If determinism is true, then  $\Box(P_0 \& L \Rightarrow P)$ . The argument that van Inwagen uses to support the falsity of determinism is as follows:

1.  $\Box(P_0 \& L \Rightarrow P)$
2.  $\Box(P_0 \Rightarrow (L \Rightarrow P))$       1; modal and sentential logic
3.  $N(P_0 \Rightarrow (L \Rightarrow P))$       2; Rule Alpha
4.  $NP_0$       Premise
5.  $N(L \Rightarrow P)$       3, 4; Rule Beta
6.  $NL$       Premise
7.  $NP$       5, 6; Rule Beta

If this argument is sound it entails that no one has or ever had any choice about anything, including what any given person does. Since this result is absurd to libertarians, determinism is shown to be false. Van Inwagen does not think anyone could dispute Rule Alpha or the two premises, so the soundness of the argument rests on the validity of Rule Beta. He does not defend Rule Beta here, but does claim that one would have no reason for being an incompatibilist if one did not accept it. Van Inwagen will have more to say about Beta in sub-section 4.2.3. Now I turn my attention to Fischer's and Ravizza's objections to Beta's relationship to incompatibilism.

#### **4.2.2. Fischer and Ravizza – Beta Jeopardized**

Though there are many Consequence arguments used by incompatibilists, van Inwagen argues that in any form they must rely on a rule of inference similar to Rule Beta. Fischer and Ravizza admit that many forms of the Consequence argument do rely on intuitions similar to those that underlie van Inwagen's Rule Beta, but they argue that it is false that one must accept Rule Beta to be an incompatibilist. To prove this, Fischer and Ravizza provide a sketch of an argument for incompatibilism that employs two principles that do not rely on Rule Beta. The two principles are the fixity of the past and the fixity of the laws of nature.

- (FP) For any action  $Y$ , agent  $S$ , and time  $T$ , if it is true that if  $S$  were to do  $Y$  at  $T$ , some fact about the past relative to  $T$  would not have been a fact, then  $S$  cannot do  $Y$  at  $T$ .
- (FL) For any action  $Y$ , and agent  $S$ , if it is true that if  $S$  were to do  $Y$ , some natural law which actually obtains would not obtain, then  $S$  cannot do  $Y$ .

Consider some act  $X$  that agent  $A$  actually refrains from doing at  $T_2$ . If determinism is true (and is taken as the thesis that a complete description of the world at  $T$  in conjunction with a complete formulation of the laws entails every subsequent truth), and  $S_1$  is the total state of the world at  $T_1$ ,

then one of the following conditionals must be true:

- (1) If  $A$  were to do  $X$  at  $T_2$ ,  $S_1$  would not have been the total state of the world at  $T_1$ .
- (2) If  $A$  were to do  $X$  at  $T_2$ , then some natural law that actually obtains would not obtain.
- (3) If  $A$  were to do  $X$  at  $T_2$ , then either  $S_1$  would not have been the total state of the world at  $T_1$ , or some natural law that actually obtains would not obtain. (244)

As stated in the excerpt, Fischer and Ravizza argue that determinism in conjunction with FP and FL entail the truth of at least one of the numbered statements. However, if (1) is true then  $A$  cannot do  $X$  at  $T_2$  because of FP, if (2) is true then  $A$  cannot do  $X$  at  $T_2$  because of FL, and if (3) is true then  $A$  cannot do  $X$  at  $T_2$  because of FP or FL. Therefore, if determinism is true then  $A$  cannot do anything other than what he does at  $T_2$ . Fischer and Ravizza argue that this incompatibilist argument does not rely on any Beta-like rule. They argue that the two fixity principles have an independent appeal, one that is not owed to any support it may or may not receive from a Beta-like rule. Therefore, “the debate over incompatibilism should not be reduced to a discussion about the validity of Beta” (244). Fischer and Ravizza do not argue that Rule Beta is invalid, only that one does not have to accept it to be an incompatibilist.

Additionally, if van Inwagen is not happy with Fischer and Ravizza’s argument, they point out that van Inwagen himself has provided two non-Beta arguments that purport to prove that determinism is false in his *An Essay on Free Will*.<sup>5</sup>

It would seem that van Inwagen has a lot of explaining to do if he would still maintain that incompatibilists must accept a Beta-like rule. I now focus my attention on his attempt at doing so.

### 4.2.3. Van Inwagen - Beta Reclaimed

Fischer and Ravizza provide an example to support their claim that incompatibilists need not rely on a rule like Beta. The example involves the principles of the fixity of the past (FP) and the fixity of the laws of nature (FL). Van Inwagen argues both that the argument form employed by Fischer and Ravizza is invalid and that the argument does not imply incompatibilism. He does so by questioning what role FP and FL play in Fischer's and Ravizza's argument.

To support his criticism, van Inwagen considers one of the two arguments he has made in *An Essay on Free Will* that Fischer and Ravizza have criticized as not relying upon a Beta-like rule. The argument is the second "Possible Worlds" argument. In it, van Inwagen uses two premises:

No one has access to a possible world in which the past is different from the actual past.

No one has access to a possible world in which the laws are different from the actual laws. (VI(b) 96)

Like the two fixity principles, the two possible world premises do not on the surface seem to rely upon the validity of Beta. However, van Inwagen questions the basis upon which we should accept the two premises. Van Inwagen provides a valid argument utilizing Beta (and Rule Alpha) that supports the second premise. Suppose that W is a world in which some actual law, L, is a false proposition.

1.  $\Box(W \text{ is actual} \Rightarrow L \text{ is false})$   
hence,
2.  $\Box(L \text{ is true} \Rightarrow \sim W \text{ is actual})$   
hence,
3.  $N(L \text{ is true} \Rightarrow (\sim W \text{ is actual}))$  [Rule Alpha]
4.  $N(L \text{ is true})$

*hence,*

5.  $N(\sim W \text{ is actual})$  [3,4 Rule Beta]

*hence,*

6. No one has access to W. (VI(b) 97)

Failing the acceptance of the validity of this argument, van Inwagen insists, one would not have any reason for accepting the second premise of the Possible Worlds argument.<sup>6</sup>

Thus, an argument that does not appear on the surface to rest upon Rule Beta, does in fact do so.

He next turns his attention to the fixity principle argument provided by Fischer and Ravizza. It would be helpful to state the argument again:

(FP) For any action Y, agent S, and time T, if it is true that if S were to do Y at T, some fact about the past relative to T would not have been a fact, then S cannot do Y at T.

(FL) For any action Y, and agent S, if it is true that if S were to do Y, some natural law which actually obtains would not obtain, then S cannot do Y.

Consider some act X that agent A actually refrains from doing at  $T_2$ . If determinism is true (and is taken as the thesis that a complete description of the world at T in conjunction with a complete formulation of the laws entails every subsequent truth), and  $S_1$  is the total state of the world at  $T_1$ , then one of the following conditionals must be true:

(1) If A were to do X at  $T_2$ ,  $S_1$  would not have been the total state of the world at  $T_1$ .

(2) If A were to do X at  $T_2$ , then some natural law that actually obtains would not obtain.

(3) If A were to do X at  $T_2$ , then either  $S_1$  would not have been the total state of the world at  $T_1$ , or some natural law that actually obtains would not obtain. (244)

The rest of their argument can be restated as follows:

(4) If (1) is true, then (via FP) A cannot do X at  $T_2$ .

(5) If (2) is true, then (via FL) A cannot do X at  $T_2$ .

(6) If (4) and (5) are true, then it follows that if (3) is true, then A cannot do X at  $T_2$ .

hence,

- (7) If determinism is true, then A cannot do anything other than what he actually does at  $T_2$ .

Van Inwagen argues that neither (1) nor (2) is entailed by the truth of determinism. The reason for this is that if  $A$  were to do  $X$  at  $T_2$ , determinism would *not* entail that a natural law that actually obtains did not obtain. This is so because it could rather be the case that the total state of the world was not  $S_1$ . Similarly, if  $A$  were to do  $X$  at  $T_2$ , determinism would not entail that  $S_1$  was not the state of the world. This is so because it could rather be the case that a natural law that actually obtains did not. Hence, neither (1) nor (2) are entailed separately by determinism. It follows from this that the conjunction of (1) and (2) is also not entailed by determinism.

At this point van Inwagen's objection becomes sketchy but I shall now try to clarify it. We can see that the falsity of (1) and (2) render (4) and (5) trivially true. From (4), (5) and (6) we can derive:

- (6') If (3) is true, then  $A$  cannot do  $X$  at  $T_2$ .

Van Inwagen would not deny the truth of (3), (6) or (6'). What he would question is what role FP and FL play in getting us from (6) to (7). (3) is a basic statement of the thesis of determinism and van Inwagen has already provided a Beta argument for indeterminism that employs a similar premise (that argument excerpted in Section 4.2.1). What *may* allow us to get from (6) to (7) is a disjunctive principle like the following:

- (FPvL) For any action  $Y$ , agent  $S$ , and time  $T$ , if it is true that if  $S$  were to do  $Y$  at  $T$ , EITHER some fact about the past relative to  $T$  would not have been a fact OR some natural law which actually obtains would not obtain, THEN  $S$  cannot do  $Y$ .<sup>7</sup>

Using FPvL we can get:

(6'') If (3) is true, then (via FPvL)  $A$  cannot do  $X$  at  $T_2$ .

Whether this would entail indeterminism (i.e., whether (3) and (6'') entail (7)) is uncertain because it is uncertain how FPvL can function within a logical proof. (Can it be cited as rule as Alpha, Beta, and Beta-Prime can be? If so, how?) How a formal proof can be constructed using FPvL is not immediately apparent. Failing this, we cannot determine whether this proof would rest (implicitly or explicitly) upon the validity of a rule like Beta. Because of this, Fischer and Ravizza have failed in their effort to show that incompatibilists do not have to rely on a Beta-like rule.

#### **4.3. Implications of Beta**

In this section, I discuss what van Inwagen takes to be the implications of Beta. Namely, that we cannot perform three types of actions: actions that we find morally reprehensible; actions that we very much want to do with no countervailing reason not to; and actions that an agent regards as the only sensible act. Since most actions fall into these three categories, we are seldom able to do otherwise. Fischer and Ravizza question whether we are actually unable to do otherwise in these three cases. They argue that there is always a relevant temporal interval in which countervailing desires can arise or in which what I refer to as existential angst can arise allowing us to do otherwise. I argue that, even though van Inwagen is inclined to agree with Fischer and Ravizza here, van Inwagen's account does not allow for the possibilities Fischer and Ravizza describe. Additionally, van Inwagen argues that even if Fischer and Ravizza are correct, it will not entail that we are able to do otherwise for an appreciably large amount of actions.

### 4.3.1. Why Beta Implies That We Are Seldom Able to Do Otherwise

Since incompatibilists must accept Beta, they must also admit that there are few, if any, occasions upon which an agent can exercise free will. Van Inwagen argues for this by showing that if Beta is valid it precludes agents from acting otherwise for three general types of acts: 1) refraining from performing acts which agents find morally reprehensible; 2) performing acts that agents very much want to do with no countervailing desire not to do it; and 3) if an agent regards an act as the one obvious thing to do or the only sensible act, the agent cannot do anything but perform that act. For the first type of act, van Inwagen offers a conditional to support his argument:

- C If *X* regards *A* as an indefensible act, given the totality of relevant information available to him, and if he has no way of getting further relevant information, and if he lacks any positive desire to do *A*, and if he sees no objection to *not* doing *A* (again, given the totality of relevant information available to him), *then X* is not going to do *A*.  
(226)

As an example, van Inwagen offers the act of lying about someone's scholarly work (call it act *A*). Van Inwagen finds such an act reprehensible and would never perform it under normal circumstances. He does not contend that, though he finds the act morally reprehensible, he would never under *any* circumstances perform this act (he would perhaps perform the act if it in turn would prevent World War III). However, in such circumstances van Inwagen would have a positive desire to do *A* which is precluded by *C*.

In an important passage, van Inwagen discusses what it would be like if *C* were to be violated. Suppose van Inwagen himself performed act *A* even though he found it to be morally reprehensible and had no positive desire to do *A* and no objection to not doing *A*. Perhaps van Inwagen had suddenly changed his mind or went berserk. The important

proviso van Inwagen makes is that *if any circumstances can be given for justifying the performance of A, a nonoccurrence of such circumstances can be included into the antecedent of C* (226). With this in mind, van Inwagen argues that C is something “very like” a necessary truth. One can no more perform an act one finds morally reprehensible (as defined by C) than one can draw a round square.

At this point, van Inwagen introduces Beta-prime, a Beta-like rule of inference whose validity he argues rests on rule Beta (as he cannot imagine anyone accepting the validity of Beta and rejecting the validity of Beta-prime):

Rule Beta-prime: From  $N x, p$  and  $Nx, (p \Rightarrow q)$  deduce  $N x, q$ . (227)

Here, “ $N x, p$ ” stands for “ $p$  and  $x$  now has no choice about whether  $p$ ”. Van Inwagen uses Beta-prime to support the conclusion that agents can not perform acts they find morally reprehensible. The argument is as follows:

$N I, I$  regard  $A$  as indefensible. (In the sense ascribed in C)  
 $N I, (I$  regard  $A$  as indefensible  $\Rightarrow I$  am not going to do  $A)$   
*hence,*  
 $N I, I$  am not going to do  $A$ .

The first premise states that I now have no choice about whether I regard  $A$  as being indefensible as described in C. He argues that I have no choice about the matter because, like most of my beliefs and attitudes, it is something I just find myself with (227). It is conceivable that I may be able to change my attitude about  $A$  over a considerable stretch of time, but not in the span of time under consideration. The second premise is a necessary truth described in C. From these two premises, Beta-prime is used to yield the conclusion that if I find an act morally indefensible then I *cannot* perform that act.

Van Inwagen investigates what it would be like to actually perform an act which one finds indefensible. It must mean that there is a future of open possibilities available to that person and in at least one of those futures he performs the indefensible act. This, van Inwagen argues, is incoherent. If I consider an act *A* indefensible, then I cannot give a description of future events that are coherently connected to the present in which I proceed to perform *A*.

Rule Beta-prime's connection with incompatibilism is given in the following argument:

- (1) If the rule Beta-prime is valid, I cannot perform an act I regard as indefensible.
- (2) If the rule Beta is valid, the rule Beta-prime is valid.
- (3) Free will is incompatible with determinism only if Beta is valid.  
*hence,*
- (4) If free will is incompatible with determinism, then I cannot perform an act I regard as indefensible. (229)

After discussing the first type of acts in which we are unable to do otherwise, van Inwagen briefly considers the remaining two. The second type of acts are those that we desire greatly to perform with no countervailing desire not to perform them. The example van Inwagen uses here is that of Nightingale in C. P. Snow's novel *Masters*. Since I am not familiar with the novel, I will create an example of my own. Consider the following scenario, which I shall refer to as the Game Show Story. Jane is a contestant on a television game show in which she competes with another contestant to answer trivia questions for cash rewards. After the host of the game show asks a question, the first contestant to buzz in with a correct answer will receive 200 points. If a contestant buzzes in with an incorrect answer, 75 points are deducted from the contestant's score. The

winner then receives the opportunity to answer one bonus trivia question for the reward of \$100,000. The game is a close one and Jane finds herself ahead by 100 points with one question remaining. She is a terrific fan of the game show and has often watched it in the company of friends at her home. While doing so, she has observed other contestants in her situation and has argued among her friends over the best strategy to employ. General agreement had been made that, in that situation, the leader should buzz in as soon as possible and attempt to answer the final question. If the leader answers correctly, the leader will win. If the leader answers incorrectly, the opponent will still have to answer correctly in order to win the game (since the leader will still be ahead after the 75 points are deducted for answering incorrectly). Jane has promised her friends that if she were in that situation she would buzz in no matter what. She *desires* very strongly two things. One, to get the chance at winning the \$100,000; and two, to not embarrass herself in front of her friends who are watching at home by not buzzing in. The host begins to ask the question and Jane is focused on her buzzer. She will not consider whether she knows the answer to the question and then buzz, but rather will save the mental processing of the question until after she has buzzed in.

In the scenario of the Game Show Story, van Inwagen would argue that Jane has no choice about whether or not she will press her buzzer after the question is read. It is open to speculation as to whether she will buzz in before her opponent, but her pressing her buzzer is inevitable. And, once she has pressed the buzzer, it is false that, given the truth of the Game Show Story, she could have done other than press the buzzer. Still, let's suppose that Jane has refrained from pressing the button. Say, she had a panic attack and fainted. Or suddenly, as the question was being asked, she came to the realization

that she was too materialistic and decided to let the opponent have a shot at winning by buzzing in first. In cases such as these van Inwagen would argue as he had earlier in the case of C. *Any coherent explanation for Jane's not pressing the button can be excluded in the example.* Say, amend the Game Show Story to include that Jane does not have a panic attack nor does she decide that she is too materialistic. With this in mind, we have the following instance of rule Beta-prime:

N Jane, the Game Show Story is true.  
N Jane, (the Game Show Story is true  $\Rightarrow$  Jane is going to press the buzzer)  
*hence,*  
N Jane, Jane is going to press the buzzer. (Or, Jane is going to press the buzzer and Jane has no choice about whether she will press the buzzer or not.) (230)

This example and argument can be expanded to include any action which agents very much desire to perform with no countervailing desires or reasons not to perform it.

Van Inwagen provides a Telephone Story as an example of the third types of acts for which agents cannot do otherwise, acts which an agent regards as the one obvious thing to do or the only sensible act. This story is similar to events we encounter in our everyday lives. Van Inwagen sits at his desk grading papers and the telephone rings. He was not expecting the phone to ring but neither was he expecting it not to ring. Without reflection or deliberation, van Inwagen puts down his pen and answers the phone. *Again, if any circumstances can be imagined for not answering the phone, the story can be amended to exclude that circumstance.* Given this, the argument is as follows:

N van Inwagen, The Telephone Story is true.  
N van Inwagen, (The Telephone Story is true  $\Rightarrow$  van Inwagen is going to answer the phone).  
*hence,*

N van Inwagen, van Inwagen is going to answer the phone. (Or, van Inwagen is going to answer the telephone and he has no choice about whether he will answer the phone.) (232)

Van Inwagen argues that since our normal, everyday situation is represented in this Telephone Story, it is not clear how many of the occasions of everyday life count as “making a choice”. When I wake in the morning I seem to be faced with the choice of getting up and going to work or staying in bed for the day. After some reflection, I see that staying in bed (say) will cause me to lose my job. Since this is unacceptable (I need the money), going to work is the only sensible thing. An apparent choice has turned, upon reflection, to be no choice at all. Van Inwagen would argue that I could not have done other than get out of bed and go to work. The relevance of this is shown in the following excerpt:

There are, therefore, few occasions in life on which--at least after a little reflection and perhaps some investigation into the fact--it isn't absolutely clear what to do. And if the above arguments are correct, then an incompatibilist should believe that on such occasions the agent cannot do anything other than the thing that seems to him to be clearly the only sensible thing. (232)

There are some cases, however, when it is not clear to an agent what to do (even when “all the facts are in” (233)). Van Inwagen lists three such cases. The first are characterized by vacillation, the second by moral struggle, and the third by indecision. The first case is quickly examined by van Inwagen. These are the “Buridan’s Ass”, “Lady-and-tiger” or “vanilla/chocolate” cases in which each available alternative is indistinguishable relative to the relevant decision making criteria. I am sent to the video store by my wife to rent *Shakespeare in Love*. I notice that there are two copies of the film in the video store and each are equally accessible to me. No amount of reflection

will allow me to conclude that either copy of the film on the video store shelf is better for me to choose than the other.

Van Inwagen later argues (235) that in cases of this sort we are not exercising free will but are instead abdicating choice in favor of an arbitrary internal decision making mechanism. In such cases we do *not* have control over the result of the arbitrary process used to determine which alternative is chosen.

The second case is characterized by moral struggle (though not every case involves morality). These are cases of duty versus inclination or cases of general policy versus momentary desire. Examples here are easily generated. Consider the case of the dieter who must decide whether to give in to the momentary desire to eat a chocolate doughnut or abstain from doing so in order to fulfill his long term goal of fitting into the army uniform he had not worn in ten years. Or, the case of a married man who is tempted by a beautiful woman. He would like to fool around with her but he would also like to be faithful to his wife.

The third case in which the agent is not sure what to do is characterized by indecision that is a result of an agent having incommensurable values. The question that confronts the agent in examples of the third case is “What sort of human being shall I be? or What sort of life shall I live?” (234). The conflict is one between a life of rational self-interest (narrowly construed to include only what is traditionally associated with “selfish” activities) versus a life of gift and sacrifice. These questions do not presuppose a set of values but rather are the questions we ask ourselves when determining what values we will accept. So, when deciding on a career the question would not be “Which profession

would enable me to have lots of girlfriends?” for this question presupposes a value. Rather, the relevant question is “What type of person would I be should I choose either?”

Since van Inwagen has argued that in cases of the first type we are not really presented with a choice, there are actually only two types of cases in which agents can exercise free will. The incompatibilist must therefore admit that there are few occasions which agents can do other than what they do. Having discussed van Inwagen’s second main thesis, I now turn to a discussion of Fischer’s and Ravizza’s objections to it.

#### **4.3.2. The Perils of Restrictive Incompatibilism**

Fischer and Ravizza address the implications van Inwagen draws from the validity of Rule Beta. To do so, they first consider the three cases in which van Inwagen had argued that agents are not able to do otherwise: the case of morally indefensible action, the case of unopposed inclination, and the case of unreflective action. Van Inwagen’s argument for each relies on either the condition van Inwagen has labeled as C (see previous sub-section) or a similar condition that parallels C. Fischer and Ravizza first discuss the case of unopposed inclination.

Van Inwagen’s argument in support of this is as follows:

- (1)  $\text{NX}$ ,  $X$  has an unopposed inclination to do  $A$ .
- (2)  $\text{NX}$ , ( $X$  has an unopposed inclination to do  $A \Rightarrow X$  is going to do  $A$ .)  
*hence,*
- (3)  $\text{N } X$ ,  $X$  is going to do  $A$ .

The second premise, Fischer and Ravizza argue, relies on the following condition C2 that parallels C:

- (C2) If  $X$  very much desires to do some act  $A$  given the totality of relevant information available to him, and if he has no way of getting further relevant information, and if he lacks any positive desire to perform

any act other than *A*, and if he sees no objection to doing *A* and refraining from doing anything else (again, given the totality of relevant information available to him), then the person is not going to do anything other than *A*. (248).

C2 can only support the second premise, however, if C2 is what Fischer and Ravizza call power necessary for the relevant agent. That is, only if C2 is true and the agent has no choice about C2 being true.

Fischer and Ravizza argue that C2 has two interpretations, each of which is inadequate for different reasons. On one interpretation, C2 is power necessary but does not support the second premise. On the other, the second premise is supported but at the cost of C2 being rendered implausible. Consider the first possible interpretation C2\*:

(C2\*) It is not possible that the following state of affairs obtain: that *X* performs an act other than *A* without having any desire to perform such an act. (249)

Fischer and Ravizza support this interpretation by arguing from the basic idea that actions are distinguished from mere events in virtue of being preceded in a suitable manner by volitions. Volitions, in turn, must also be based in a suitable manner on desires. It follows from this that it would be impossible for an agent to perform an action without having some desire to do so. However, this interpretation does not imply the second premise:

(2)  $\neg X, (X \text{ has an unopposed inclination to do } A \Rightarrow X \text{ is going to do } A.)$

Fischer and Ravizza provide an example in which C2\* is true and (2) is false. This example depends upon there being what they term a “relevant temporal interval” (249)

between the moment  $X$  has the unopposed inclination and the moment  $A$  would be performed during which  $X$  is able to generate an alternate desire. To reconsider the Game Show Story in the previous section, we can imagine that Jane has a strong desire to press the buzzer but during “some temporal interval” between that moment and the moment she would have pressed the buzzer, Jane develops a desire to not press the button.<sup>8</sup> In this case, C2\* would be true because if she had not developed an opposing desire she would have pressed the buzzer. (2) is false because at some moment Jane had an unopposed inclination to press the buzzer yet she did not in fact press the buzzer (because she subsequently developed an opposing desire before she could act on the original desire).

Now consider the second interpretation of C2:

(C2\*\*) If  $X$  does not desire to do other than  $A$ ,  $X$  cannot do other than  $A$ . (249)

Under this interpretation, C2 does indeed support (2). However, it does so at the price of plausibility. Fischer and Ravizza argue that an agent having no desire to perform an action does not preclude him from performing the action because the agent could always generate relevant desires to motivate him to perform the action. Fischer and Ravizza provide an example of this desire generation in the following excerpt:

Just about anybody can summon up the worry that he is not free to do otherwise. That is, one can worry that, despite the pervasive intuitive feeling that frequently we have genuine freedom to do various things, we do not in fact have such freedom. (250)

It is just such a worry that can cause Jane to decide to not press her button after she had developed a previously unopposed inclination to do so. At one point, Fischer and

Ravizza would argue, Jane had an unopposed inclination to press her button. This entails that at that moment she had no desire to do other than A. However, subsequently, existential angst (if you will) caused her to generate a desire to prove that she was not a pawn of fate by refraining from pressing the buzzer.<sup>9</sup> So it seems that in cases of unopposed inclination we *are* able to act otherwise.

Fischer and Ravizza use the above argument as a blueprint for their attack against van Inwagen's remaining two cases, that of morally indefensible action and the case of unreflective action. They provide two possible interpretations for the relevant "C" claim and shows that if these interpretations are true then they do not imply the second premise of the relevant argument. If the second premise is held to be true, then the relevant "C" claim is shown to be implausible. They conclude in the former that agents are able to perform morally indefensible actions (just to show that they are free to perform the action, as it were). In the latter, they conclude that agents are able to refrain from performing unreflective actions. Fischer's and Ravizza's emphasis in this section of the paper is on morally indefensible actions, so I shall turn my attention there.

Fischer and Ravizza give two examples of agents who perform acts they consider morally indefensible. The first is that of St. Augustine stealing pears as a youth. Though Augustine did see the theft as having some desirable consequences (namely, the "thrill of having partners in sin"), he still performed the action that (it can be argued) he found to be morally indefensible. They opine that Augustine could have been motivated by a perverse sort of freedom or power -- "a freedom to ignore the Good" (254).

The second example is taken from the writings of Dostoevsky. In *Crime and Punishment*, the character Raskalnikov contemplates killing and robbing an old

pawnbroker even though he knows that it is morally indefensible. Despite his moral aversion, he is able to commit the crime. In this case, Raskalnikov seems to commit the act “precisely to see if he *can* do it: he wants to discover if he has the power to ignore moral prohibitions; he wants to know if he is free to do the morally indefensible.” (255). These two examples are used to underscore the position that agents can do otherwise in cases of morally indefensible actions.

Fischer and Ravizza have argued that van Inwagen’s three cases in which agents are unable to do otherwise are flawed because there is always the chance for the agent to develop a relevant counter-desire. Van Inwagen responds to this by arguing that even if this is true, it would result in few additional cases in which an agent could have done otherwise.

#### **4.3.3. When the Will is Not Free**

Van Inwagen responds to Fischer’s and Ravizza’s criticism that the validity of Beta does not entail that we are able to act otherwise as seldom as van Inwagen claims. Originally, van Inwagen had argued that there are at least three general types of cases in which we are unable to do otherwise: cases of morally indefensible action, unopposed inclination, and of unreflective action. Fischer and Ravizza argued that in those three cases there can be a “relevant temporal interval” during which desires can arise (perhaps via what I have termed “existential angst”) which would enable the agent to do otherwise. In the current paper under consideration (VI(b)), van Inwagen agrees with the basics of Fischer’s and Ravizza’s criticism and amends his original position. However, van Inwagen argues that this does not entail that we are able to do otherwise as often as

Fischer and Ravizza would like. We can see why van Inwagen concludes this by examining the cases of morally indefensible actions.

Consider two examples that Fischer and Ravizza provided - that of St. Augustine and that of Raskalnikov. Each performs an action they consider indefensible. Each has done so in an attempt to show that they were not pawns of fate (perhaps). Augustine stole the pears to show that he was free from the Good, and Raskalnikov committed his murder to show that he was beyond good and evil.<sup>10</sup> Whether these examples work or not depends upon when Augustine and Raskalnikov developed their desires to prove their freedom. In order for it to be a counterexample to van Inwagen's position, there must have been some point at which they held their respective potential actions to be morally indefensible and had no such desires to prove their freedom. If this were not so, the actions would not count as ones they found morally indefensible as outlined by C

C If *X* regards *A* as an indefensible act, given the totality of relevant information available to him, and if he has no way of getting further relevant information, and if he lacks any positive desire to do *A*, and if he sees no objection to *not* doing *A* (again, given the totality of relevant information available to him), *then X* is not going to do *A*.

because C precludes the possession of positive desires to perform the indefensible act. Therefore, the desires must have arisen, to use Fischer's and Ravizza's phrase, during some "relevant temporal interval" between the moment that, say, Augustine determined that stealing pears was an indefensible act and the moment he actually stole the pears.

Van Inwagen is willing to concede that his argument is flawed (or at least not obviously true). However, he is quick to point out that this does not imply that agents are able to perform a significantly higher number of free actions for two reasons. First, van

Inwagen argues that cases when “existential angst” arises in the consideration of morally indefensible actions are rare. Second, even in cases where an agent has “existential angst” when considering an indefensible action, rarer still are the instances in which the agent then proceeds to commit the indefensible action. Therefore, van Inwagen’s overall claim that free actions are very rare still holds.

Oddly, van Inwagen does not take issue with the role of the “relevant temporal interval” in Fischer’s and Ravizza’s arguments. As I have mentioned earlier in a footnote, van Inwagen has set up the conditionals (the C statements) in such a manner as to exclude *ex hypothesi* the strategy that Fischer and Ravizza employ. Van Inwagen argues that C is something like a necessary truth and instructs that if *any* circumstances can be given for justifying the performance of an indefensible act then the nonoccurrence of such circumstances can be included into the antecedent of C (226). So, the antecedent of C can be amended as follows:

If *X* regards *A* as an indefensible act, given the totality of relevant information available to him, and if he has no way of getting further relevant information, and if he lacks any positive desire to do *A*, and if he sees no objection to *not* doing *A* (again, given the totality of relevant information available to him), *and if there is no relevant temporal interval between the moment X regards A as an indefensible act and the moment at which the final choice about whether A is performed such that X can generate a desire to perform A, and if X does not have a case of existential angst*, then *X* is not going to do *A*.

In addition to this defense, van Inwagen could also argue that cases of existential angst violate the condition that the agent has *no way of getting further relevant information* concerning *A*. I do not immediately see how a future desire (in whatever

form it may take) to perform *A* is not considered relevant information concerning the indefensibility of *A*.

I am not yet entirely convinced by my move. Fischer and Ravizza could argue that though the case of existential angst is relevant to the argument in general and to the actual performance of the indefensible act in particular, it is *not* relevant to the indefensibility of the action itself and is thus not precluded by C. However, even if my objection can be answered by Fischer and Ravizza because the existential angst is not relevant to the indefensibility of the act, I think the problem of existential angst *is* a part of the *totality of relevant information* available to the agent. Because of this, Fischer's and Ravizza's objection that we are actually able to do otherwise in cases of morally indefensible acts (as defined by van Inwagen) may not hold.

But let us entertain the "relevant temporal interval" employed by Fischer and Ravizza further. Fischer and Ravizza seem to be relying upon something like the following statement (in the case of morally reprehensible actions):

At any moment in time, no matter how small the moment, an agent is able (via existential angst) to generate a desire to perform an action she finds morally reprehensible.<sup>11</sup>

If this is true, and if the possibility of a "relevant temporal interval" is excluded via the relevant conditional statement, then it will be true that an agent cannot perform an action he finds morally reprehensible *only* at the exact moment that he succeeds in not performing the morally reprehensible action. At every second of time previous to that, the "relevant temporal interval" objection would come into play and the agent would be able to generate a perverse desire to perform the act he finds reprehensible. Even if this

is admitted to be the case, however, agents would not be free to do otherwise appreciably more than they are under van Inwagen's because of the arguments given previously in this section.<sup>12</sup> This can be shown utilizing the Game Show Story used in section 4.3.1. (This example is one of an action that one desires greatly to perform with no countervailing desire not to perform it rather than that of a morally reprehensible action, but it will still illustrate the relevant point.)<sup>13</sup>

Jane stands posed at the podium with her finger upon her buzzer. She has a very strong desire to press the buzzer and no desire to not press the buzzer. The final trivia question will be asked in four seconds. During those four seconds, Jane can develop a perverse desire to not press her buzzer. Though it can be admitted that Jane has this power, it will seldom be the case that she actually does so. Additionally, once the desire not to press the buzzer arises, fewer still will be the cases in which Jane will heed the new desire and fail to press the buzzer.

To recap, I do not agree that Fischer and Ravizza have shown that agents can perform morally indefensible actions (as they are defined by van Inwagen). However, even if Fischer's and Ravizza's criticism is correct (and van Inwagen certainly thinks that it is), the number of free actions an agent could perform would not significantly increase.

#### **4.4. Beta and Moral Responsibility – The Classical Tradition**

In this section I consider the question of whether or not van Inwagen's account of moral responsibility resides in the classical tradition. Though he has argued that we are seldom able to do otherwise, van Inwagen does not think that this results in an inability to hold agents responsible in a great amount of cases. This is so because of something that Fischer and Ravizza refer to as the tracing principle. This principle states that we can

hold an agent responsible even in cases in which he could not have done otherwise provided there was a time when the agent could have foreseen the future action and could have arranged events such that it did not occur. Fischer and Ravizza question van Inwagen's account because it seems that most of our actions are determined by our character that is in turn determined at an early age by our upbringing and our surroundings. Holding agents responsible in such cases seems to go against our every day ideas of responsibility and accountability.

#### **4.4.1. When Can We Be Held Responsible? – Van Inwagen**

Van Inwagen discusses what implications his conclusions have for questions of moral blame. Van Inwagen uses “drunk driver” cases to show that the classical tradition should not be committed to the thesis that an agent can be held accountable for a state of affairs only if he either intentionally brought that state about (or could have refrained from bringing it about) or if that agent foresaw that the state would obtain unless he prevented it and that he was able to prevent it. It can be argued that a drunk driver who swerves into oncoming traffic neither set out to do so nor could he have prevented it (due to his intoxication). Though it is true that he could have refrained from getting drunk, the actual outcome was not foreseen by the agent. Obviously, however, we would still hold the drunk driver responsible for his action. Van Inwagen claims that this relationship between blame and free will can be expressed as follows:

An agent cannot be blamed for a state of affairs unless there was a time at which he could so have arranged matters that that state of affairs not obtain. (236)<sup>14</sup>

This does not imply that we are responsible for all of the consequences of our actions but rather only those that are in some sense foreseeable. To return to the case of the drunk driver, he knew before he started drinking that he would be driving later. It was foreseeable that his chances of causing an accident would be greatly increased. Therefore, he is a candidate for responsibility. Suppose that the individual in question were an alcoholic and on that particular night he could no more refrain from drinking and driving than a rock could refrain from falling to Earth when dropped. Then there must have been some point after the time he took his first drink that he could see that he was on the road to alcoholism (and all the “evils” associated with it, like drunk driving). Because of this, he is held responsible for the accident. I will see if van Inwagen’s position is unassailable by examining objections raised by Fischer and Ravizza.

#### **4.4.2. Questions About Responsibility – Fischer and Ravizza**

Fischer and Ravizza then turn to showing that restrictive incompatibilists like van Inwagen cannot provide a satisfying theory of moral accountability while still remaining within the classical tradition. The classical tradition holds that there is an intimate connection between free will and moral responsibility such that if there were no free will (and no one was ever able to do otherwise) then there would be no moral responsibility. This suggests that any state of affairs for which an agent can be held responsible must be able to be “traced” back to a prior free action performed by the agent (259). Fischer and Ravizza refer to this as the “tracing principle” and do not think that it bodes well for restrictivists who wish to remain within the classical tradition.

Since the restrictivist claims that we are free in at most three types of situations (Buridan cases, cases in which duty conflicts with inclination, and situations of conflict

between incommensurable values), he is committed to showing that any and all states of affairs for which an agent can be held responsible must be traceable back to one of those three types of situations. The restrictivist can hold that the conflict situations characterized by two of the three situations (the latter two situations) are the ones through which an agent's character is formed. Following from this, the agent can still be held responsible for actions that are produced by his character. As an example, consider the case of the drunk driver who is an alcoholic as described in the previous section. He is held responsible for the accident because it resulted from the character he has built up (by becoming an alcoholic). Fischer and Ravizza spot a flaw in this reasoning, however.

Much of our character results from the habituation we receive in early life, and these portions of our character don't seem to be necessarily connected with situations of conflict between duty, inclinations, or incommensurable values. (260)

Fischer's and Ravizza's point here is that an agent cannot be held responsible for actions that result from his character if that character was formed via "habituation we receive early in life". Since presumably most of an agent's character is formed through such habituation, it seems that the cases in which an agent can be held responsible are even rarer than the restrictivist claims. To illustrate his point, Fischer and Ravizza provide the example of a young woman named Betty.

Betty was raised in a rural community where patriotism and American pride is prevalent. Growing up, Betty has this pride and patriotism instilled within her and a conflict never arises which could cause her to call her patriotism into question. While traveling in a foreign country, Betty is approached by a foreign agent. The agent asks Betty to betray her country in return for some monetary profit. Betty finds treason to be

morally indefensible and immediately turns down the agent. In this case, Betty's action can be said to have flowed from her character. However, the restrictivist must claim that Betty's action is not worthy of praise because it cannot be traced back to a free action for which Betty was responsible. Fischer and Ravizza argue that this is absurd because such a conclusion "runs directly counter to our actual practices of holding people responsible" (261). Because of this, the restrictivist position is outside that of the classical tradition. Fischer and Ravizza have argue that the restrictivist position does not rest within the classical tradition because the restrictivist position entails consequences in the realm of moral responsibility that are not acceptable within the classical tradition. As I show in the next sub-section, van Inwagen does not agree with this assessment.

#### **4.4.3. Moral Responsibility – Van Inwagen**

Van Inwagen responds to Fischer's and Ravizza's criticism that states that restrictivism entails that we can seldom if ever be held morally accountable for what we have done. Van Inwagen has earlier argued that an agent can be held morally accountable for an action that he could not have at the time refrained from performing if it can be shown that the agent is accountable for having the inability. This is exemplified by the case of the drunk driver who is an alcoholic. Though he could not have refrained from getting drunk in that particular instance, he can still be held responsible because (presumably) he at one point made a free choice which ultimately resulted in his becoming an alcoholic. Fischer and Ravizza would agree with this example but argue still that restrictivism entails that agents can seldom be held responsible.

As may be recalled, Fischer and Ravizza reach this conclusion after they consider the following:

Much of our character results from the habituation we receive in early life, and these portions of our character don't seem to be necessarily connected with situations of conflict between duty, inclinations, or incommensurable values. (260)

Also recall Fischer's and Ravizza's example of patriotic Betty. Our actual practices of assigning moral responsibility would require us to praise Betty for declining to commit an act of treason. However, the restrictivist would be forced to argue that Betty's action is not morally praiseworthy because it was an expression of her character and her character was not a result of free choice by Betty. Fischer and Ravizza argue that this conclusion is absurd and, as such, the restrictivist is not within the classical tradition in the debate concerning free will.

Van Inwagen takes issue with Fischer's and Ravizza's claim that our actual practices of assigning moral responsibility would require us to praise Betty for her action because of a certain asymmetry inherent in those practices. This asymmetry is "between bad and good or between approval and disapproval" (VI(b) 108). Cases in which moral responsibility is ascribed are typically those that involve states which ought not obtain. In other words, the typical cases of moral responsibility are those in which blame is attributed to an agent rather than a positive credit. Because of this, in order for Fischer and Ravizza to reach their conclusion they must provide a counterexample that involves blame being attributed to an agent rather than praise. Unless and until they are able to do so, van Inwagen is not excluded from the classical tradition.

I do not consider Fischer's and Ravizza's appeal to our everyday practices of holding agents morally responsible to be a legitimate move on their part. Even if the asymmetry that van Inwagen appeals to was nonexistent, Fischer and Ravizza must still contend with the idea that our everyday practices of holding agents morally responsible are not solely based upon whether an agent is actually responsible for the action. Rather, there are also forward looking consequentialist reasons for praising Betty's action regardless of whether Betty herself is worthy of such praise. It is not immediately obvious that philosophers in the classical tradition would not agree that Betty is unworthy of praise if they are told the story in its entirety. At the very least, there would not be anything resembling *uniform agreement* that her action is praiseworthy relative to her history. What they may agree upon, however, is that praising Betty's action would have beneficial effects because such praise is likely to influence others to act patriotically.

Additionally, Peter Strawson argues in "Freedom and Resentment"<sup>15</sup> that discovering the thesis of determinism to be true should not deter us from holding reactive attitudes towards one another. Among these reactive attitudes are ascriptions of moral responsibility. Even if Betty's character was determined by her upbringing, Strawson would argue, we should still hold Betty responsible because that is the type of creatures we are. We can not help but praise her for performing an action of which we approve because of the "human commitment to ordinary inter-personal relationships."<sup>16</sup> Though I do not agree with Strawson, his view shows that Fischer's and Ravizza's example of Betty does not serve to place van Inwagen outside the classical tradition because, even within that tradition, there is no wholesale agreement over whether or not we should hold Betty responsible for her act.

#### **4.5. Concluding Remarks**

Even if libertarian theories of freedom and agency like Kane's can be made intelligible, they face the problem of not being as free as they may like. Even if we possess libertarian freedom, it may also be true that in almost every instance we are not able to do other than what we do. At the beginning of this chapter I considered the problem that, even if we possess libertarian freedom, it may also be true that in almost every instance we are not able to do other than what we do. Van Inwagen has argued that this is the case, but that it does not imply that we are only seldom responsible for our actions. He has done so by appealing to Beta, a rule of inference. I have given a representation of the philosophical positions held by Peter van Inwagen and John Martin Fischer and Mark Ravizza in their respective papers and I have demonstrated that Fischer and Ravizza have failed to show that incompatibilist philosophers are *not* required to rely on a rule of inference such as van Inwagen's Beta. I also have defended van Inwagen's position that Beta implies that we are seldom if ever able to do other than what we do and have shown that Beta does not further imply that we can seldom if ever be held morally responsible for our actions. Though van Inwagen's position was originally seen by Fischer and Ravizza to be setting unwanted and unneeded limits to libertarian freedom because of the limited ability of agents to do otherwise, van Inwagen has successfully shown that this should not worry libertarians because under certain conditions agents can be held responsible even in those cases where it is true that agents could not have done otherwise.

## **4.6 End Notes**

<sup>1</sup>In O'Connor (1996), 219-238.

<sup>2</sup>In O'Connor (1996), 239-269.

<sup>3</sup>Van Inwagen (1994).

<sup>4</sup>All subsequent page references in this chapter will refer to the O'Connor volume unless otherwise noted.

<sup>5</sup> This point is made in footnote 8 on 242. The two arguments made by van Inwagen are the First Formal Argument and the second "Possible Worlds" argument. Van Inwagen reveals how Beta supports these two arguments in VI(b).

<sup>6</sup> Van Inwagen argues that a similar argument can be made for the First Formal Argument that he employs in *An Essay on Free Will*.

<sup>7</sup> Van Inwagen does not explicitly mention a principle such as this, nor does his response to Fischer and Ravizza run exactly as I have stated here, but it does parallel the form of his objection on VI(b), 98-99.

<sup>8</sup>It would seem that this type of objection is eliminated *ex hypothesi* by how van Inwagen has set up the argument. Van Inwagen would have us add to the antecedent of C2 that there would be no "relevant temporal interval" during which X can change his mind. However, he risks weakening his claim to the point of being a trivial truth. I discuss this possible defense of van Inwagen in section 4.3.3.

<sup>9</sup>A similar response can be made here as in the previous footnote – see section 4.3.3.

<sup>10</sup>I took this description from footnote 11 (VI(b) 112).

<sup>11</sup> This statement is my creation and is not taken from Fischer's and Ravizza's article.

<sup>12</sup> Namely, that even though the possibility to generate the opposing desire exists, very seldom will the desire actually be generated. Further, in cases where the desire is generated it will very seldom be acted upon.

<sup>13</sup> Fischer and Ravizza discuss an objection along similar lines (251). The objection is the complaint that they have simply missed van Inwagen's point. The objection is credited in a footnote to Sara Buss, Nancy Schaubert, and Eleonore Stump.

<sup>14</sup> Of course it is always true that we could have arranged matters such that a particular state of affairs not obtain. It is the foreseeability requirement that saves this from being a trivial truth.

<sup>15</sup> In Watson, 59-80.

<sup>16</sup> Watson, 68.

## 5. Conclusion

I started this work with the intention of examining and evaluating the indeterminist solution to the problem of free will and free agency. To do so I examined a recent work of Robert Kane, one of the foremost indeterminist philosophers currently engaged in the discussion concerning the problem of free will. Kane has offered a unique account that attempts to place libertarian philosophy upon the same footing as compatibilists regarding its scientific plausibility. However, Kane's account falls prey to criticisms that I have lodged in chapter two as well as criticisms lodged by other philosophers discussed in chapter three. First, Kane does not provide an acceptable account of the significance of libertarian free will because he does not recognize the importance of the connection between the significance question and the intelligibility question. Second, Kane is not able to account for the problem of moral luck regarding how indeterminate efforts of will are resolved. Third, rather than strengthening the indeterminist position, Kane's use of folk psychology serves to make compatibilist accounts of agency more intuitive. Because of these, Kane has not been successful in providing an intelligible account of freedom.

For an indeterminist to meaningfully employ indeterminism in gaining freedom, he must answer the four objections I have discussed in chapter three. Unlike the three objections in the previous paragraph, these four objections apply to any and all indeterminist accounts of freedom and not specifically to Kane's. An acceptable indeterminist account must have at least four features. First and most importantly it must be able to use indeterminism in such a way that it is meaningful in the problem of free will. Second, he must be able to show how actions performed by an agent are actions for

which the agent can be held responsible rather than actions that simply happen. Third, he must satisfactorily explain why the indeterminism involved must be of the absolute random variety as opposed to the epistemically random. Fourth, he must be able to explain how actions that are the result of an indeterminate process can be considered actions that are rational.

Even if an indeterminist philosopher were able to provide acceptable solutions to these problems, he would be faced with the additional problems examined in chapter four of this work. As van Inwagen has argued, even if determinism were false it would not result in a significant amount of cases in which agents are able to do other than what they in fact do. This inability to do otherwise in a significant number of cases would seem to indicate that we could only seldom be held responsible for our actions. At least with respect to Fischer's and Ravizza's objections, van Inwagen has adequately argued that this is not the case. Though we are unable to do otherwise in a significant number of cases, we can still be held responsible for actions that are determined by our character.

Despite van Inwagen's successful refutation of Fischer's and Ravizza's objections, all is not well with his account. This is because I am not convinced that agents should be held responsible for actions that flow from their character, even if they are responsible for their characters. I devote the remainder of this work to sketching a few additional problems for the indeterminist, including the problem of actions that are determined by character.

### **5.1 Further Problems for Indeterminists**

In this section I briefly develop several additional objections to libertarians that rely upon indeterminism. First, I question whether we should hold agents responsible for

actions that flow from their characters even on those occasions in which the agent is responsible for shaping his character. Second, I use the concept of overdetermination to provide a rough sketch of a world in which universal determinism is true yet agents can still be held responsible for their actions. In doing the latter, I question whether conditions such as Kane's AP condition are relevant to the problem of free will.

### **5.1.1 Responsibility and Actions Determined by Character**

Both Kane and van Inwagen have discussed the topic of whether or not an agent can be held responsible for an action that flows from his character. In the Martin Luther example Kane argued that agents could be held responsible for actions that flow from their character provided they are responsible (via a previous self-forming action or SFA) for possessing their current character. Luther's past actions (including several SFA's, of course) formed his character such that Luther both could not have done other than break from the Church of Rome and yet is responsible for his action. That this is so is demonstrated by Kane's UR condition for freedom of the will. Recall that sub-condition U of UR indicates when an agent can be held responsible for an action that flows from his character:

(U) for every X and Y (where X and Y represent occurrences of events and/or states) if the agent is personally responsible for X, and if Y is an *arche* (or sufficient ground or cause or explanation) for X, then the agent must also be personally responsible for Y. (Kane (1998), 35)

In this context, Luther's break with the Church of Rome is X while the totality of his past actions that determined his character (and that act as sufficient grounds for X) is Y. Because Luther is responsible for Y, he is also responsible for X.

Van Inwagen argues similarly using the example of a drunk driver. Once drunk, it seems we cannot hold the agent responsible for his decision to get behind the wheel and place his life and the lives of others in danger. Van Inwagen argues that this is not the case because of the “tracing principle.”<sup>1</sup> It is as follows:

An agent cannot be blamed for a state of affairs unless there was a time at which he could so have arranged matters that that state of affairs not obtain. (O’Connor, 236)

In the case of the drunk driver, we can hold him responsible for his act provided there was a time in the past at which point he could have foreseen that his consuming alcohol might get him drunk and, once drunk, he might stupidly go for a drive. Thus an agent’s ability to foresee the future action plays a part in his ability to arrange matters differently. If he could not foresee the future action, he could not have taken precautions. This tracing principle works similarly to Kane’s sub-condition U. In this case, driving drunk serves as *X* while *Y* is represented by the agent’s decision to continue to drink alcoholic beverages. Since *Y* is sufficient grounds for *X* (*X* can be traced to *Y*), and since the agent is responsible for *Y*, the agent is also responsible for *X*. Before continuing with this discussion, it would be helpful to make a few distinctions.

### **5.1.2 ADC, AR and ANR**

To aid in discussion of this topic, I refer to actions that flow from an agent’s character as ADC (for actions determined by character). Further, I distinguish between situations where the agent is responsible for his character (and hence responsible for the ADC) and situations where the agent is not responsible for his character (and hence not responsible for the ADC). In the former cases, as typified by Kane’s use of the Martin Luther example and van Inwagen’s drunk driver example, the agent is responsible for the

ADC via the agent's responsibility for the formation of the his own character. These types of actions I refer to as AR (for agent responsible). AR actions are ADCs for which the agent can be held responsible.

The latter type of ADC includes cases in which an agent's character is determined solely by his environment or his genes. An example of this can be found in the film *Ravenous*. The protagonist in the film is an officer in the United States armed forces during the Spanish-American War who, to escape being killed by the enemy, feigns death on the battlefield. Thinking he is dead, the Mexican soldiers place him in a pile of dead bodies that include several of the officer's posthumous subordinates. American soldiers attack the Mexican position and the officer is able to sneak out of the pile and kill the commanders of the Mexican forces. Because of his experience amongst the blood drenched dead soldiers, he is later unable to partake of the very rare prime rib served at a banquet in his honor. In this example, the officer's inability to eat the prime rib is an ADC but not one for which the officer is responsible. I refer to this type of ADC as ANR (for agent not responsible). An ANR is an ADC for which the agent is not responsible.

It is important to note that AR and ANR refer to the sufficient grounds for the agent's character, not the action that flows from the character. Thus, as I argue in the next section, there may be examples of an ADC that is an AR but for which the agent is not responsible. In these cases, the agent is responsible for his character (hence, AR), but he is not responsible for action that is determined by his character.

### **5.1.3 The Perils of Character Building**

Generally, when AR actions are discussed there is an obvious and direct link between the actions that the agent has performed to build his character and the resulting

ADC. In the Martin Luther case, Luther performs various character building actions that are of the same sort as his breaking with the Church of Rome. His choosing to act piously and by what he views to be the correct religious standards determine an action in which he similarly acts piously and by those same religious standards. Here the transfer of responsibility from the original acts to the later ADC is clear. However, not all AR actions result in an ADC that is of a similar nature.

Consider the case of Swagger, a man of strict religious stature who has worked diligently to become a perfect Christian.<sup>2</sup> He has an idea of what a perfect Christian is and sets his will towards performing actions that he thinks will help him achieve it. His efforts to shape his own character are successful, so successful that we can say that he is the sole cause of his resulting character.<sup>3</sup> This makes his resulting character of the AR variety. Swagger's character then determines several actions that are of a similar variety. He does not debate whether to go to church on Sunday or to go fishing, or if he does deliberate it is a mere pretence. He will go to church. Swagger will not steal money from the collection plate. He will not reject his religion and become a Communist. However, Swagger is not omniscient. His efforts to become the ideal Christian man go astray for two reasons.

First, his conception of the perfect Christian could be unattainable, say, because it is not within the power of any human being to act so perfectly. Because of this, his character shaping has had the unforeseen side effect of placing enormous amounts of stress and pressure upon him. The stresses eat at him so much that he eventually performs a very un-Christian act with a prostitute in a motel on Airline Highway. This action is an ADC, and the character that determined the action was AR. Using Kane's *X*

and *Y*, Swagger's character building actions, *Y*, served as sufficient grounds for the un-Christian act, *X*. Because Swagger is responsible for *Y*, he is thus also responsible for *X*. Thus, we should hold Swagger responsible for the ADC.

I find this unacceptable, at least within the limits of this example. *Ex hypothesi*, once he had shaped his character Swagger could not have done other than perform the un-Christian act. Again, *ex hypothesi*, he could not have foreseen that this un-Christian act would result from his character shaping. Because of his inability to have foreseen the consequences his responsibility for the un-Christian action can be questioned. This poses a problem for Kane's condition UR because he does not account for this possibility. This is less of a problem for van Inwagen, because according to the tracing principle responsibility can be attributed only if the agent could have arranged things so that the later action did not occur. In order to arrange things differently, the agent must have been able to foresee that the later action could occur. In the case of the drunk driver, that he could have foreseen that his drinking could lead to his driving drunk. In the present case, that Swagger could have foreseen that his good intentions could lead to an un-Christian action. Since his eventual action was not foreseeable in this version of the Swagger example, van Inwagen's account can allow for Swagger not to be held responsible for the ADC even though Swagger is responsible for his character.

Though van Inwagen's tracing principle is capable of handling the Swagger example, the question of foreseeability can pose problems for van Inwagen. Recall from chapter four that van Inwagen has argued that we are free to do otherwise on very few occasions and also that this does not entail that we can seldom be held responsible for our actions. The latter is so because of the tracing principle. Agents are unable to do

otherwise most of the time. But, so long as their inability can be traced back to those actions for which they are responsible and regarding which it is true that they could have done otherwise, their responsibility for the earlier action is enough to make them candidates for being held responsible for the latter, determined action. We can take the next step and actually hold the agents responsible so long as the determined action was foreseeable by the agent. Van Inwagen must give a clear account of the criteria for foreseeability in order for his philosophy to be useful. It is certainly true that Swagger *did* not foresee the future consequence of his action, but that does not imply that he *could* not.

Suppose we learn something new about Swagger. In addition to being very religious, he is also a practicing psychologist with an intimate knowledge of the frailties of the human mind. Because of this, he has learned what happens to the mind when it is put under too much pressure. In short, he is aware that it can cause agents to do other than what they normally do. Additionally, he has treated several patients who have tried to live the ideal Christian life but have cracked under the tremendous pressure resulting in performance of un-Christian actions. With this in mind, does Swagger qualify for being able to foresee that his road may lead to an un-Christian act? He is certainly aware that it is a possibility. Because of the many patients who have tried and failed, he may even consider it likely. Still, he feels duty bound to follow the Christian ideal and ultimately breaks under the pressure. Is his awareness that it was possible or likely enough to make the resulting un-Christian action foreseeable? If so, then van Inwagen would (wrongly, I think) hold him responsible for the ADC. If not, then it may be that (imperfect beings that we are) we may have adequate knowledge to foresee very little. We may be able to,

say, foresee immediate consequences of our actions, but the longer the amount of time between the ADC and the actions that make the ADC an AR, the less likely it is that the agent could have foreseen what action would have resulted.

A second reason why Swagger's attempt to become an ideal Christian can go wrong is, even if it is granted that the ideal is attainable, the path to reaching it may be indeterminate. It may be that being a good Christian is not something that you can strive to become, but rather only stumble upon while trying to do something else. In the process of trying to reach his ideal, Swagger may unintentionally become something less savory all together. The road to hell is said to be paved with good intentions. Because of this, Swagger has no rationally necessary means that is associated with his adopted end. Again, because we are not omniscient beings, Swagger does not know this. After much deliberation, he sets down and follows a path and begins to shape his character. After much time has passed he recognizes that he is not making progress towards his end and has instead ended up shaping his character in a way that is not to his liking. His efforts have resulted in shaping a certain character within him, a character that he does not necessarily want. Once having done so, should we hold Swagger responsible for actions that flow from this character? In this case I have not specified that Swagger could have foreseen that this might occur. Because of this van Inwagen's tracing principle may not apply. I will now construct another case in which the tracing principle does apply.

Bubba is an earnest young man who is disgusted by the political corruption that runs rampant in his state legislature. He is so dismayed that he decides to run for the office of state senator for his district. He manages to prevail in the election and once in office attempts to fight corruption wherever it appears. It is a lonely battle, however, and

the enemy is strong. He finds that the only way he can make progress is to trade votes. In exchange for supporting various bills that Bubba puts forth, a corrupt senator asks that Bubba support another bill. Bubba recognizes that the bill that he will be forced to vote serves to fatten the pockets of both Bubba and the corrupt state senator. Bubba initially is not concerned with money but does recognize that it has seduced many politicians before him. If he accepts the bargain, he may well eventually become what he hates. However, he may also make real progress and gain enough allies to rid the state legislature of corruption. With the intention of doing the latter, unfortunately, his decision to trade votes leads to the former. Little by little, he is seduced by the money until his character is as corrupt as any of the others. At each point, Bubba has attempted to do the right thing (or attempted not letting doing the wrong thing adversely affect his future character). At each point, he fails. In short, he ends up with a character that he does not want. His efforts have missed the mark. Though he did not want to become a corrupt politician, the possibility was foreseeable by Bubba. Hence, according to van Inwagen's tracing principle we should hold Bubba responsible for the latest act of corruption that is determined by his character. Though he may constantly work to change his character and may actually eventually succeed in regaining his honesty, he is responsible for the unwanted actions that flow from his unwanted character. I find this to be counterintuitive. In cases like this, the agent seems to be acting against his will and should not be held responsible for the resulting actions even though his character was determined previously by his will.

Each of these cases is one for which an ADC action that was AR did not result in an action for which we should necessarily hold an agent responsible. What I leave

unaddressed is the question of how often cases such as these arise. I think that without proper deliberation there is always the possibility that we will perform an action that we either do not foresee, or one that was foreseeable but not likely, or one that was foreseeable and likely but an action that the agent did not want to result from his careful character shaping actions. Because of this, there is no immediate reason for believing they do not occur with frequency.

#### **5.1.4 Compatibilism**

I now want to address several problems associated with the compatibilist position. I realize that compatibilists have for the most part not been discussed directly in this work, and I only do so here in an attempt to describe a further problem for incompatibilist philosophers. This problem deals with what Kane calls condition AP (for alternative possibilities), the condition that in order for an action to be considered free it must be true that the agent could have (in some sense of this phrase) done otherwise. I question what relevance the libertarian interpretation of alternate possibilities has to the problem of free will.<sup>4</sup>

Libertarians typically paint a rather bleak picture of a world that is governed by universal determinism. I would like here to sketch an alternate, at the very least less bleak, picture of that world in which agents may be held responsible for their actions. I do so by utilizing the concept of overdetermination. Before doing so, it would be fruitful to clarify what is meant by overdetermination.

#### **5.1.5 Overdetermination**

Overdetermination occurs when there is more than one cause ( $C_1, C_2... C_n$ ) that is sufficient for bringing about a particular effect ( $E$ ) and more than one of the causes occur

simultaneously. For example, there are several causes that are sufficient to bring about the effect of my death. In this case, let  $C_1$  be a murderer shooting me in the heart using a rifle and let  $C_2$  be my accidentally ingesting a deadly poison that is mislabeled as medicine (but which takes some time to work). Both  $C_1$  and  $C_2$  are each alone sufficient to cause  $E$ , my death. Since they are each individually sufficient they are, of course, jointly sufficient to cause  $E$ . Suppose that I accidentally ingest the poison but at the exact moment before I expire I am also shot in the heart by the murderer. Sadly, I am dead. But which cause acted to bring about  $E$ ? In this case, had  $C_1$  not occurred then  $C_2$  would have brought about  $E$ . Had  $C_2$  not occurred,  $C_1$  would have brought about  $E$ . Either way (in this example), my death was inevitable. Both  $C_1$  and  $C_2$  determine  $E$  and, hence,  $E$  is said to be overdetermined. Overdetermination, however, does not serve to diminish responsibility. Even though it is true that had the murderer not fired his shot I would still be dead, we would not relieve him of the responsibility for my murder. Similarly, even though I would have died anyway had I not accidentally ingested the poison, my loved ones can hold the company that mislabeled the medicine responsible for my death.

### **5.1.6 Overdetermination and Compatibilism**

A compatibilist philosopher can appeal to overdetermination in an attempt to explain why conditions like AP (conditions that require that we “could have done otherwise”) are not important in ascribing responsibility to agents. Instead of arguing over how alternative possibilities should be interpreted, compatibilist philosophers can just reject it altogether. The compatibilist can agree that if determinism is true it is also true that we can never under any circumstances do other than what we in fact do because our actions are caused by the conjunction of events that occurred before we were born

and the laws of nature. In this way determinism serves as  $C_1$  for effect  $E$  (any particular action we perform). However, this does not close the book on the question of responsibility. Just because our actions have one set of sufficient conditions that actually obtain does not preclude there being another set of sufficient conditions,  $C_2$ , that also serve to simultaneously determine our actions. In this case  $C_2$  can be the (perhaps folk psychological) decision making process agents employ when they determine which action to perform. Because of the overdetermination at work, agents can be held responsible for their actions

There is an asymmetry between the two examples that must be acknowledged. In the case of the murderer and my accidentally poisoning myself, if either the  $C_1$  or the  $C_2$  did not occur I would still end up dead. I am not sure how in the latter case we can sensibly talk about  $C_2$  occurring in the absence of  $C_1$ . The question is whether the agent's decision to act did play any causal role since there is *no* possibility of  $C_1$  *not* determining any resulting action. In the determined world,  $C_1$  is always present. Additionally, in the first case the sufficient causes were not casually related to each other in the way the latter case is held to be. My choosing to perform an act is casually related to determinism such that the forces of determinism cause my desire. This asymmetry is important and must be sufficiently explained. Though such an explanation would go beyond the scope of this present work, I do offer a suggestion.

Even if it is granted that they should be treated the same, we are still faced with the problem of why, in the case of agents deciding to act, we should break the causal chain at that point and not extend it back to before the birth of the agent. In this case  $C_2$  collapses into  $C_1$ . Some explanation must be given for why the causal chain should be

considered broken before it extends in such a manner. But what would this explanation be? In truth, I do not know. It may be that we have to simply acknowledge and accept that for various reasons (our biological complexity, our capability of higher thought, etc.) we should treat each other differently than we treat other (in this context possibly lesser) creatures for which questions of responsibility do not arise. Though it is true that the past and natural laws determine our actions, it is also true that we are the ones that perform these actions and that they are actions that we intend to perform. To the extent the two coincide, our actions being determined by the past and the laws of nature and also being actions that we want or desire to perform, overdetermination results and we can be held responsible for the action. This use of overdetermination is hardly satisfactory and must be discussed and examined further. It was not my intent to do so here. Rather, my intention is to suggest it as a possibility.

If overdetermination can be used in allowing for agents to be responsible for actions in a determined world, then the compatibilist can safely reject conditions such as AP. In the next sub-section, I briefly discuss what would result from doing just that.

### **5.1.7 Could Not Have Done Otherwise**

“Genie,” says Max while firmly holding the bottle that serves as its home, “when I was in high school I desperately wanted to ask Janice to the prom but I did not do so. I blame all of my failures as a man since then upon that decision. If only I had the courage, if only I could have asked, my life would have been vastly different. Vastly better without this doubt gnawing at me at every moment. Genie, I wish that I had the chance to do it all over again. I want the chance to ask her to the prom again.”

The genie slowly spreads his arms apart and then brings his hands together in a thunderous clap. “Your wish is granted.”

Max braces himself, not knowing what to expect. How radically different would his life be now that he does not have to live with the memory of this mistake any longer? Moments pass, but nothing happens. “Genie? Did it work? I don’t feel any differently,” Max hesitantly asks.

“Yes, master. I have turned back time to the exact moment of your decision. Unfortunately, you again decided not to ask her to the prom. Did you expect otherwise?”

I conclude with a discussion of what exactly is lost when conditions that require alternative possibilities are abandoned. Usually, when we wish for chance to do something over again, like Max, we mean that we want to somehow have known back then what we know now and use this information to alter the choices we have made. Max does not simply want the chance to ask Janice to the prom again. The genie granted him that and Max was not satisfied. What Max wanted was to change the past such that his younger self somehow had knowledge of his future failures and used that knowledge to motivate him to ask Janice to the prom. However, this would not have been the same situation. Max would not be doing it all over again, for if he were (given the truth of determinism) it would turn out just as it did the first time. Even were he to make the same wish countless times the results would be the same. This is because there must be some *difference* in the past in order to *make* a difference in his younger self’s decision. What would it look like if he were to decide to ask her to the prom? His younger self would not choose to ask Janice to the prom unless there was, say, a reason to ask her that he did not originally consider or if, say, the genie magically altered the younger Max’s character. Again, it would not be a case of doing it all over again because the two situations would be different.

Though this has been framed from the point of view of a compatibilist, the incompatibilist who relies upon indeterminism to gain freedom should also question whether or not conditions such as AP are relevant to discussions of freedom. The short story of Max given above would undoubtedly be quite different for these libertarians. It

may be that Max's wish results in his younger self deciding to ask Janice to the prom. The younger Max, intending not to ask Janice to the prom, suddenly does the opposite. This would result in the familiar problem raised by Double's objection from rational explanation in chapter three.

I am not arguing that either side is better off here, but rather that the problems associated with alternate possibilities are great and the reward for the indeterminist is small. If alternative possibilities are allowed, we have actions that could have happened otherwise. The indeterminist is then faced with a similar objection that is typically lodged against the compatibilist. In the case of the compatibilist, the challenge is to provide an intelligible account of freedom and agency that results in determined actions equating with free choices. In the case of the indeterminist, the challenge is to provide an intelligible account of freedom and agency that results in indetermined actions equating with free choices. Though the problems are on the surface different, I think they stem from the same larger problem of (and I agree with Kane here) how materialistic accounts of the self can allow for responsible action by agents, whether determined or not.

## **5.2 Concluding Remarks**

I have devoted the bulk of this concluding chapter to sketching possible problems for the libertarian philosopher who relies solely upon indeterminism to obtain freedom. First, I have questioned the idea that we should hold agents responsible for actions that flow from their character as long as they are somehow responsible for having their character. In doing so, I have examined conditions given by Kane and van Inwagen for holding agents responsible in these cases. I have found Kane's UR condition unacceptable, and I have argued that van Inwagen must provide acceptable criteria for

foreseeability in order for his version to work. Next, I have attempted to paint a more favorable picture of a world governed by determinism using the concept of overdetermination. This was done in an attempt to show that conditions such as Kane's AP (alternative possibilities) condition are not necessary for free will. Additionally, I have argued that we do not lose anything meaningful when we discard conditions like AP because given the chance to do everything over again we would proceed to do everything in the exact same way we have done so previously. Each of these problems requires more development and investigation, however, before they could be considered threats to the incompatibilist position.

### **5.3 End Notes**

<sup>1</sup> It is actually Fischer and Ravizza that refer to this as the "tracing principle." Van Inwagen does not refer to it in this manner, but I will do so for the sake of simplicity.

<sup>2</sup> I don't pretend to know what an ideal Christian is nor have any expertise in philosophy of religion in any form. If the use of the ideal Christian is distasteful, the ideal of a perfectly moral creature can be substituted in its stead.

<sup>3</sup> It is doubtful that a situation such as this could exist because there are so many factors that influence us in ways that we cannot predict. I set this aside in order to make the example simpler.

<sup>4</sup> Of course, I have already discussed extensively cases where agents could not have done otherwise yet are still responsible for their actions. These are cases in which the agent is responsible for his inability to do otherwise and are safely ignored in this section.

## References

- Berofsky, Bernard (ed.). 1966. *Free Will and Determinism*. New York: Harper & Row.
- Dennett, Daniel. 1978. *Brainstorms*. Cambridge, Mass.: Bradford Books.
- Dennett, Daniel. 1984. *Elbow Room*. Cambridge, Mass.: MIT Press.
- Dworkin, Gerald. 1986. Review of Dennett, *Elbow Room*. *Ethics* 96: 423-5.
- Frankfurt, Harry. 1969. "Alternative Possibilities and Moral Responsibility." *Journal of Philosophy* 66: 829-39.
- Honderich, Ted. 1988. *A Theory of Determinism*. 2 Vols. Oxford: Clarendon Press.
- Kane, Robert. 1985. *Free Will and Values*. Albany: SUNY Press.
- Kane, Robert. 1998. *The Significance of Free Will*. New York: Oxford University Press.
- Lewis, David. 1981. *Philosophical Papers*. 2 Vols. Oxford: Oxford University Press.
- Molesworth, W. (ed.). 1962. *The English Works of Thomas Hobbes*. Vol 5. London: Scientia Aalen.
- Morgenbesser, Sidney, and J. H. Walsh (eds.). 1962. *Free Will*. Englewood Cliffs, N.J.: Prentice-Hall
- Nagel, Thomas. 1979. *Mortal Questions*. Cambridge: Cambridge University Press.
- O'Connor, Timothy (ed.). 1996. *Agents, Causes and Events*. Oxford: Oxford University Press.
- Schopenhauer, Arthur. 1960. *Essay on the Freedom of the Will*. Trans. by K. Kolenda. Indianapolis: Bobbs-Merrill.
- Strawson, Galen. 1994. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75: 5-24.
- van Inwagen, Peter. 1983. *An Essay on Free Will*. Oxford: Oxford University Press.
- van Inwagen, Peter. 1994. "When the Will is Not Free." *Philosophical Studies* 75: 95-113.

Waller, Bruce. 1988. "Free Will Gone Out of Control: A Critical Study of R. Kane's *Free Will and Values*." *Behaviourism* 16:149-67.

Watson, Gary (ed.). 1982. *Free Will*. Oxford: Oxford University Press.

## VITA

Troy Dwayne Fassbender was born in New Orleans, Louisiana, on July 21, 1970. He graduated from Alfred Bonnabel High School in Metairie, Louisiana, in 1988. He earned a certification in Paralegal Studies at the University of New Orleans and worked as paralegal for approximately ten years before deciding to return to university. He earned a Bachelor of Arts in Philosophy and Religious Studies from Louisiana State University in 1999 and will obtain a Master of Arts in Philosophy and Religious Studies there in May of 2002. Troy met Joyce, a graduate student in the entomology department and now his wife, while they were both in-line skating around the campus of Louisiana State University. They have each been offered admittance into the Graduate Center of the City University of New York. They will move there in August of 2002.