

**AGENCY AND PERSONS:
HOW WE BECOME WHO AND WHAT WE ARE**

ALESSANDRO FIORELLO

Thesis submitted to the University of Ottawa
in partial Fulfillment of the requirements for the
Doctor of Philosophy in Philosophy

Department of Philosophy
Faculty of Arts
University of Ottawa

© Alessandro Fiorello, Ottawa, Canada, 2025

Abstract:

This doctoral dissertation articulates a libertarian theory of free will and moral responsibility alongside a narrative view of personal identity. In this dissertation, I build upon and expand Robert Kane's libertarian theory to create a mitigation strategy for dealing with a perennial problem for libertarian theories of freedom and moral responsibility: the problem of luck. I argue that Kane's basic idea of self-forming actions or SFAs can be built upon to show how the luck objection can be undercut. I argue that SFAs take place within a larger narrative structure and that when we make SFAs we are also engaged in what Kane calls value experiments. I construct a theory of narrative personal identity and put it to work to show how the problem of luck is mitigated and that SFAs are not just a matter of luck.

Keywords:

Libertarianism, Free Will, Moral Responsibility, Robert Kane, Personal Identity, Luck Objection

Acknowledgements:

Many people have supported me during this project. I want to thank the University of Ottawa and the Ontario Graduate Scholarship for their financial support of the thesis. My advisor Dr. Andrew Sneddon read countless drafts and provided invaluable feedback that shaped the project for the better. I want to thank my wife Hilary, who has been by my side every step of the way of this process, and I would have never taken the plunge to go to graduate school without her encouragement. She is my co-author to the story of my life. My mother always asked about and supported the project. Alla mia papà, mi manchi e vorrei tu fossi qui con me. So che ne saresti orgoglioso.

Table of Contents

1. Introduction.....	3
2. Libertarianism	15
3. Compatibilism	62
4. Incompatibilism	100
5. Authorship and Agency	150
6. Libertarianism and Luck	196
7. Conclusion	223
Bibliography	228

Chapter 1

Introduction

There is a disputation [that will continue] till mankind is raised from the dead, between the necessitarians and the partisans of free will.¹

-Jalalu'ddin Rumi

1. An Ancient Philosophical Problem

The problem of free will of which Rumi speaks is one of philosophy's most ancient and persistent conundrums. From the ancient Greeks and Romans to our contemporary period the topic of free will is one of philosophy's most perennial problems.²

When I was first introduced to this issue as an undergraduate it was presented like so: when we go about our daily lives, we are faced with choices we have to make. These include the trivial and the significant, from choosing between chicken and tomato soup for lunch to choosing a life partner. Now, it seems the concept of choice implies being able to pick either A (chicken soup) or B (tomato soup).

A potential problem is posed by the idea of determinism. Determinism is the thesis that if we take any moment in the history of the universe and combine it with the laws of nature what follows is *necessitated* by prior causes.³ Suppose I am faced with making a choice between going to a Toronto Maple Leaf's game or Shakespeare's *Romeo and Juliet*. If determinism is true, then

¹ Jalāl al-Dīn Rumi, *Rumi, Poet and Mystic*. Translated and edited by Reynold Alleyne Nicholson. (London: Allen & Unwin, 1956), 77.

² Anthony Kenny, *A New History of Western Philosophy* (New York: Oxford University Press, 2010).

³ Robert Kane, "The Problem of Free Will: A Libertarian Perspective." In *Do We Have Free Will?: A Debate*, edited by Robert Kane and Carolina Sartorio, 3-67 (New York: Routledge, 2022).; Peter Van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1983).; Helen Steward. *A Metaphysics for Freedom* (Oxford: Oxford University Press, 2012).

the laws of nature in conjunction with the past necessitate whether I go to the hockey game or the Shakespeare play. There is no “open future” since whatever I do, the necessitating causal chain of the laws of nature entails there is only ever exactly one possible future. This is contrary to the way I feel when making a choice as it seems to me that at least sometimes I feel I could have done otherwise. The problem of free will comes into sharp focus for me when I hold in contrast my feeling about my actions and the possibly deterministic causal chains that produce my action.⁴ On the one hand, it follows from the truth of determinism that I can never do otherwise, while, on the other hand, my subjective experience seems to include exactly this ability that determinism rules out. In a deterministic universe all events including our actions must occur exactly as they do.

The understanding of free will that is in tension with determinism is one where free will entails the power or ability to have done otherwise. Our understanding of free will as an ability to do otherwise seems critical for moral responsibility. It is *prima facie* false that an agent could be held blameworthy or praiseworthy if their actions were the necessary entailment of the past and the laws of nature. Moral responsibility appears on its face to entail that agents have the ability to control their actions. Control is ambiguous and as I discuss later on there are different kinds of control available to agents. I contend there is one specific sense of control that entails agents have the ability to perform or refrain from performing an act and this is at least sometimes tied to being moral responsible. This claim seems *prima facie* plausible, but a full defence will come over the course of the discussion. This specific sense of control is in tension with determinism because determinism implies that our actions are necessitated by the past and the laws of nature.

⁴ There is a debate in the experimental philosophy literature on free will over whether our feelings about free choices are aligned with compatibilism or incompatibilism. A classic discussion of the issue is Shaun Nichols and Joshua Knobe, “Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions,” *Noûs* 41, no. 4 (2007): 663–85.

If true, this would mean that when we act, we could not refrain. So, we would not have the control required for moral responsibility. Consider the following case: Smith is deliberating whether to kill the mayor of Ottawa. On the *prima facie* understanding of control, Smith has control for his action if he has the ability to kill the mayor or refrain from killing the mayor. This ability to perform or refrain from performing his action is a form of being able to do otherwise which would be ruled out by determinism.

This line of thought seemed quite intuitive to me; but I was then introduced to arguments that tried to persuade me that any worry about the incompatibility of either free will or moral responsibility with determinism must be unfounded, grounded in confusion about the meaning of freedom or adhering to an incoherent and absurd ideal. These philosophers, known as compatibilists, argue that our freedom and moral responsibility remain thoroughly unscathed by determinism. I'll call this "generic compatibilism." Their arguments, which we will see in chapter 3, are interesting but they ultimately left me unconvinced; and so here I am attempting to argue that we should indeed be worried about freedom and responsibility if determinism turns out to be true. The contrary of compatibilism—incompatibilism—seems to me to be closer to the truth.

2. Free Will: Ultimacy, Authorship, and Control

I propose an understanding of free will that involves in a qualified way an ability to have done otherwise. Also, it requires much more besides than merely being able to do otherwise. The following is a brief definition of free will to be used, elaborated and defended throughout the dissertation: free will is the power of agents to be the authors of their own life stories. It must be stated and indeed emphasized that I employ "author" as a metaphor; I am not saying everyone is a literal author and that free will and moral responsibility involve literal writing. Being the author

of one's life entails one is the author of oneself, one's own ends, goals, and purposes.

Authorship involves doing otherwise but in a qualified way. The kind of free will I am therefore going to develop is a libertarian conception. Libertarianism is a doctrine that states free will is incompatible with the truth of determinism and that human beings sometimes act freely.

I will have much more to say on about this view in chapter two but for the present introduction let us begin with what I think is a fundamental motivation for the theory. In the last section I discussed the worry about responsibility concerning control. Control comes in different forms, and libertarianism asserts that a specific kind of control entailing an ability to act or refrain from acting is crucial for responsibility and is ruled out by determinism because we cannot refrain from determined actions. The kind of control where we can perform or refrain from performing an act is important because, I think, it is the kind of thing we need in order to say that our actions are, "up to us." What it means to say that an action is "up to us" is ambiguous. This idea of action being "up to us" is a key feature of libertarianism but could be interpreted along compatibilist lines. The significance of this idea will be seen in the discussion of libertarianism in chapter two. We can also see the ambiguity of the idea of an action being "up to us" through the two interpretations contemporary theorists take as described by Kevin Timpe: sourcehood and alternative possibilities.⁵ The sourcehood approach requires that an action be traceable in its causal history to the agent in a particular way, while the alternative possibilities or *leeway* view

⁵ Kevin Timpe, *Free Will: Sourcehood and Its Alternatives* (London: Continuum Books, 2008).

posits that an agent must be able to do otherwise.⁶ Both of these approaches can be articulated in compatibilist or incompatibilist terms.⁷

The understanding of “up to us” that will be fully articulated and defended in chapter four is in terms of sourcehood. This might come across as surprising given the discussion so far, and while being able to do otherwise has a role to play on my account, sourcehood plays the more fundamental role in my account because I believe we don’t always need alternative possibilities to say we are the source of our identities. This ties into a third idea related to sourcehood: ultimacy. Incompatibilists care about an action being “up to us” because if an act is up to us we are the ultimate source of it. Some incompatibilists believe that agents could fulfill the requirement of ultimacy without ever being able to do otherwise. Others believe the present sense of “up to us” is to be understood precisely in terms of alternative possibilities. In my view, ultimacy and alternative possibilities are both important, but I maintain that alternative possibilities are necessary only on certain occasions.⁸

Being the ultimate source of our actions means that agents are the, “buck-stopping originator(s)” of our actions.⁹ I think we are the ultimate source of our actions because we are the authors of our actions and ourselves. Ultimate sourcehood in terms of authorship will be explored in chapter five but here I’ll introduce the idea in terms of tracing the production of our

⁶ Derk Pereboom, *Living Without Free Will* (Cambridge: Cambridge University Press, 2001), 3.

⁷ For an example of a compatibilist account that maintains alternative possibilities as necessary for free will, see, Peter Forrest, “Backward Causation in Defence of Free Will,” *Mind* 94, no. 374 (1985): 210–17. For an example of sourcehood compatibilism, see, Harry Frankfurt, “Freedom of the Will and the Concept of a Person,” In *The Importance of What We Care About* (Cambridge: Cambridge University Press, 1988), 11-25.

⁸ “Rethinking Free Will: New Perspectives on an Ancient Problem.” In *The Oxford Handbook of Free Will*, edited by Robert Kane, 381-404. Second Edition. New York: Oxford University Press, 2011.

⁹ Galen Strawson, *Freedom and Belief* (New York: Oxford University Press, 1986), 26.

action back to something inside us like our wills. Consider an agent who is presently intoxicated and is thus unable to control their actions in that they are unable to refrain from performing an act. While such agents are not presently in control, their actions can be traced back to when they had the control to drink and not to drink. A full discussion and defence of ultimacy would be out of place here; in keeping with the spirit of the introduction I will present the meaning of “ultimacy” to be used throughout the thesis. Ultimacy in the intended sense entails that the source of an act is within an agent. This is a rough and imprecise statement, and a more precise meaning will be unravelled throughout the thesis. What I want to emphasize now to avoid misunderstanding is that ultimacy does not entail that agents have total control or that agency requires agents be isolated from their environment. We do not need to be able to do otherwise for every action we perform for it to be performed of our own free will. Our actions could be determined while being consistent with ultimacy if we were the ultimate source of our wills. The requirements and nuances of ultimacy and how we shape our wills will be explored in chapter four. This points to a distinction to be explored in chapter two about a difference between freedom of action and freedom of will. Freedom of action is compatible with determinism but in a specific sense to be discussed and defended in the thesis freedom of will is incompatible with determinism. The sense of freedom of will that is incompatible with determinism points to the relevance of ultimacy.

What ultimacy would require is that an action could not be traced to things agents do not control like their genetic inheritance, environment, upbringing, God’s foreknowledge etc. Ultimacy permits genetics and upbringing to have an influence, but it does not permit these factors to be necessitating. I will discuss tracing and its role in free will in the next chapter. For example, “Why did Jones leave the room?” “Because his friend died last week.” Jones’ leaving

the room involves things he obviously does not control but his leaving the room is not determined by antecedent causes. That his friend died last week works alongside Jones to make him leave the room. Jones “works alongside” antecedent causes in the sense that they influence him but whether he acts or refrains from acting remains up to him. Jones does not act independently of the antecedent causal inputs, but his action is not necessitated. While this is backward-looking, I hold that ultimacy is not only concerned with the past but also the future. Ultimacy is concerned with determinism being true because it implies that the past implies only one possible future. This means that the past not only necessitates who one is but also who one is to become and at no point in one’s entire lifetime is one ever able to become different than who one is.

I view ultimate sourcehood as important for control because, if an action can be traced to an agent and then further to necessitating causes, then the agent will not be able refrain from performing that act. Now, we can of course say in a compatibilist sense Jones made a free choice because there was no coercion or compulsion; and in the absence of any force Jones controlled his decision because he acted on his desire. But this is not enough from an incompatibilist libertarian perspective. We need not only to be free to act but to be free to will or as I prefer to say to author what it is we want to act on. Chapters 3, 4, and 5 deal with the arguments for the requirement for ultimate sourcehood and why compatibilism falls short on exactly this score.¹⁰

¹⁰ I view this conception of free will and moral responsibility as integral to various other aspects of human life: love, creativity, desert, and the meaning of life. Obviously, compatibilists will likely claim none of these are exclusive to an incompatibilist understanding of free will, but I think they are integral to the libertarian’s concern about ultimacy regarding determinism. For instance, it surely matters to persons that they are the source of their creative talents and accomplishments. It matters that Beethoven was the source of his great ninth symphony and that it was not necessitated by deterministic causal forces.

This understanding of free will has sustained continued critique from various flanks. From philosophy, compatibilists have continued to argue that what libertarians want must be either confused, undesirable, or unnecessary. From the sciences, the latest advances in social psychology, physics and neuroscience have attempted to paint an empirical picture of human beings with no place for libertarian free will.¹¹

These philosophers and scientists, I think, are mistaken in dismissing libertarianism as a plausible option. I will not be delving into the scientific aspect of the free will problem in the subsequent discussion of my view so I will not say anything further on this front.¹² Some philosophers in the last half of the previous century thought free will to be a dead issue not worth our attention.¹³ They are wrong. There is so much more to explore and argue about on this topic with many unpaved roads to discover.

3. Authors and Stories

The view defined in the previous section intertwines free will with another long-standing philosophical issue: personal identity. Indeed, the problem of free will touches so many facets of human inquiry but perhaps, I think, none more fundamental than the question of who and what

¹¹ Benjamin Libet, “Do We Have Free Will?,” in *The Oxford Handbook of Free Will*, ed. Robert Kane, First Edition. (New York: Oxford University Press, 2001), 551-558.; , Daniel M. Wegner, *The Illusion of Conscious Will* (Cambridge, Massachusetts: MIT Press, 2002). ; Sam Harris, *Free Will* (New York: Free Press, 2012).

¹² To say something on this, an example of where I think they are wrong to dismiss libertarian free will is the experiments conducted by Benjamin Libet in the citation above. He conducted experiments where he claimed to have predicted what subjects have decided before they consciously became aware of their decisions. I think the claim that free will does not exist to not be supported by the experiment. For a discussion which offers an argument for Libet’s experiments failing to undermine libertarian free will, see Alfred Mele, *Free: Why Science Hasn’t Disproved Free Will*. (New York: Oxford University Press, 2014).

¹³ Aldoph Grunbaum, “Free Will and Laws of Human Behavior,” *American Philosophical Quarterly* 8, no. 4 (1971): 299–317.

we are. The view of free will proposed argues that as authors persons are ultimate creators and it is through the ultimate creation of one's story that one also creates oneself.

As we will see this view of persons as authors has come to prominence beginning at the end of the twentieth century and firmly entrenching itself in the literature in the contemporary field. This view believes we have to distinguish different ways we can talk about persons and focuses on how to attribute characteristics to individual persons. The view of free will to be explored and defended is one that believes a necessary condition for attributing these characteristics as one's own is that one has a hand in shaping them through freely willed choices. It is in specific moments where we make certain choices that we write or author our stories and ourselves. The association I draw between free will and personal identity is one that, I think, is an untravelled path that ought to be explored.

4. Structure of the Thesis

I begin in chapter 2 by outlining libertarianism. I argue that Robert Kane's libertarian theory is where libertarians should look to for their theorizing. Kane's basic argument is that free will requires a specific kind of action between incommensurable alternatives which he calls "self-forming actions". For Kane, free will is historical in the sense that in order for our wills to be free we have to have had a hand in shaping them.¹⁴ The proper guiding questions in his account are firstly, how did we come to be the kind of persons that we are and secondly, did we become these people by our own making or are we entirely the products of forces we have no control over. This line of thought seems to me to be the right track to pursue a libertarian theory of freedom and moral responsibility.

¹⁴ Robert Kane, *The Significance of Free Will* (Oxford: Oxford University Press, 1996).

Chapter 3 engages with compatibilist arguments that try to establish that determinism poses no threat to responsibility practices or our freedom. I survey classical, twentieth century and contemporary theories and argue why they are all lacking. What fuels my intuitions that compatibilism is insufficient as a complete account of moral responsibility and free will is that I view compatibilist responsibility and freedom as an incomplete picture. To be sure, compatibilist accounts offer valuable varieties of freedom but they do not accurately describe all the significant kinds of freedom agents can possess. I think the freedom to (figuratively) author oneself is one such freedom.

Chapter 4 is where I argue for the connection between personal identity and free will. I argue that free will and moral responsibility have a sourcehood criterion which is here understood as authorship. I examine John Martin Fischer's critique of ultimate sourcehood through his rebuttal of Galen Strawson's Basic Argument. Fischer is correct in his assertion that Strawson's argument is very demanding but misses, I think, a key insight into moral responsibility as well as free will: ultimacy. Persons are the ultimate creators and thus the ultimate source of their lives because they are authors: They create themselves through the (metaphorical) writing of their stories. Without this I claim persons could not have what Kane calls ultimate responsibility for their wills or their stories.¹⁵

In chapter 5 I offer an account of authorship as a form of agency. Building on the work done by philosophers in debates about personal identity I develop my own narrative view of the self called "the narrative self-formation view." (NSFV hereafter). The NSFV draws attention to an important part of personhood: authorship. Persons are (metaphorically) self-authoring beings

¹⁵ Kane, *The Significance of Free Will*, 39.

who engage in what John Stuart Mill coined “experiments in living.” The narrative self-formation view (NSFV) is inspired by the work of Kane and Marya Schechtman. The major objection the theory deals with in chapter 5 has to do with unity. As the theory makes an analogy between the lives of persons and literary narratives we have to contend with what I will call the “problem of complexity”. Literal persons are complex in the sense that they take on many different practical roles in their lives; their careers, friends, relationships, and so on are all kinds of practical roles with their own identities. The problem of complexity put simply is that on my view we require a unity to our life stories; one single identity that unifies all the different identities we are. But why is this necessary? Is this even possible? So the worry goes people have different identities and the different identities in one’s life need no unification into one single identity or story.

Chapter 6 deals with a perennial problem for libertarian theories: the problem of luck. Now as Nadine Elzein notes the problem comes in more than just one form. I am going to focus on one specific form of the problem from luck.¹⁶ This branch of the problem of luck asks us to consider explanations for undetermined actions and why they happened as they did as opposed to how they could have gone. The problem demands that actions must be explainable in a contrastive fashion in order to attribute control to agents over undetermined actions. What is a contrastive explanation? A general and noncontroversial definition of contrastive explanations is that they are explanations where the explanandum (the fact to be explained) is why x rather than y. X is the fact and y is the foil that is contrasted with x. An example from the literature constructed by

¹⁶ Nadine Elzein, “The Demand for Contrastive Explanations,” *Philosophical Studies* 176, no. 5 (2019): 1326. Elzein discusses alternatives instances off the problem of luck such as “the disappearing agent” objection as well as claims that indeterminism has to be random and uncaused.

Fred Dretske helps to elaborate this definition.¹⁷ The General turns the key in the ignition of his car and the car explodes. A contrastive explanation asks why turning the key caused the explosion rather than starting the car.¹⁸ We can ask the same question for undetermined actions. Why did Jane choose to go to law school rather than go to graduate school? The worry as far as indeterminism goes is that we are unable to explain why Jane chooses one option rather than another so it is just an inexplicable happening. The most influential form of this species of argument is put forward by Al Mele, and so I will zoom in on his articulation of the objection.¹⁹

I argue my NSFV is able to respond to this concern. Mele's formulation is specific in his demand for contrastive explanations. Mele argues that we need explanations which contrastively explain an action in terms of, "the agents' powers, capacities, states of mind, moral character, and the like that explains the difference in outcome."²⁰ The kind of explanation that Mele appears to demand is one that highlight features of an agent and then contrastively explains an action in agential terms. Kane believes there can be no contrastive explanations for SFAs because contrastive explanations are deterministic and SFAs are undetermined.²¹ When we ask why X happened rather than Y whatever the answer it seems to necessitate what happens or not happens. I will argue that the relationship between an agent and her narrative retains her status as the ultimate source of her choice.

¹⁷ Fred I. Dretske, "Contrastive Statements," *The Philosophical Review* 81, no. 4 (1972): 411–37.

¹⁸ Alessandro Fiorello, "Free Will, Values, and Narrative Selfhood," *Philosophia* 48, no. 1 (2020): 7.

¹⁹ Alfred R. Mele, "Kane, Luck, and the Significance of Free Will," *Philosophical Explorations* 2, no. 2 (1999): 96–104.; Alfred R. Mele, "Libertarianism, Luck, and Control," *Pacific Philosophical Quarterly* 86, no.3 (August 2005): 381–407.; Alfred R. Mele, *Free Will and Luck* (Oxford: Oxford University Press, 2006).

²⁰ Mele, "Kane, Luck, and the Significance of Free Will," 99.

²¹ Kane, *The Significance of Free Will*.

Chapter 7 concludes the dissertation with a discussion of some implications of my account in other related areas of philosophy. I draw out the implications my view has on value theory. A consequence of my theory is that I see what is valuable is not knowable *a priori* and only knowable through Millian experiments in living which involve the free choices of agents. I discuss what Kane calls “value empiricism” and the contrasting view “value rationalism”. I ultimately conclude that human agents are free agents and the nature of their freedom is what I have tried to illuminate.

I have quite the task ahead of me in articulating a coherent and persuasive account of libertarian free will. The difficulty is compounded by the fact that libertarianism has not grown any more persuasive in philosophy while compatibilism has ascended to be the favoured view. I therefore have quite the steep uphill climb ahead of me so I need to start climbing.

Chapter 2

Libertarianism

1. Introduction

The topic I aim to discuss is libertarian free will. My goal is to establish that Robert Kane's framework is the way forward for libertarianism. I hope to demonstrate that the Kanean understanding of self-formation is crucial to a strong and robust libertarian view. My discussion in this chapter will proceed first with a brief overview of what libertarianism is generally speaking before moving on to the particulars. The first section briefly outlines what negative and positive theses libertarianism endorses. I will further break down the dominant approaches in the field. I move on in section three to provide the details of Kane's framework. I offer my assessment and what I take to be its strength and weaknesses. Taken as a whole the strengths outweigh the weaknesses in Kane's approach. In section four, I delve into the literature to evaluate the competing libertarian theories on offer. My goal is to establish that as a whole Kanean libertarianism is *the* framework libertarians should adopt in making sense of an incompatibilist notion of free will.

2. Freedom, Responsibility and Determinism

Libertarianism as a doctrine can be summarized in the following schema:

- (1) Free will and moral responsibility are incompatible with the truth of causal determinism.
- (2) Either human agents have free will or determinism is true

This disjunction has two possible answers:

(3) Determinism is true and human agents lack free will.

(4) Agents do have free will and determinism is false.²²

The first thesis characterizes libertarianism as an incompatibilist theory. It states that it is not logically possible that determinism is true and that individuals are free and morally responsible for their actions. Incompatibilism is divided between libertarians and hard incompatibilists (I discuss some examples in the next chapter) and libertarianism affirms free will exists and is committed to (4). Libertarianism therefore believes that no one can be a bearer of free will or morally responsible in a fully deterministic universe. Of course, the truth of determinism is an empirical matter that currently remains unsettled.²³ In light of this my goal is to provide a sensible and important incompatibilist type of freedom.

In chapter one, I introduce the importance of the idea of an action being “up to me”. We can now outline in terms of libertarian conditions what it means for an action to be ‘up to me.’ On the sourcehood variety of incompatibilism, being the source of one’s actions is incompatible with determinism as one’s actions would be traceable to factors beyond one’s control. Only if we can trace the source of her action to something in her and nothing further does this view allow an action to be ‘up to me.’ Since tracing does a significant amount of work here it needs clarifying. Tracing has been discussed in relation to moral responsibility by John Martin Fischer and Neal Tognazzini.²⁴ Consider the following case from Fischer and Tognazzini: Suppose that Dan gets

²² Laura Waddell Ekstrom, *Free Will: A Philosophical Study* (Boulder, Colorado: Westview Press, 2000), 1-18.; Laura Waddell Ekstrom, “Introduction,” in *Agency and Responsibility: Essays on the Metaphysics of Freedom*, ed. Laura Waddell Ekstrom (Boulder: Westview Press, 2001), 2-3.; Christopher Evan Franklin, *A Minimal Libertarianism: Free Will and the Promise of Reduction* (New York: Oxford University Press, 2018), 2.

²³ Mark Balaguer, *Free Will as an Open Scientific Problem* (Cambridge, Massachusetts: MIT Press, 2010), 131-169.

²⁴ John Martin Fischer and Neal A Tognazzini, “The Truth about Tracing,” *Noûs* 43, no. 3 (2009): 531–56.; Tracing is a subtopic in its own right in the literature and I am not able for reasons of scope to go into detail. For more

drunk at his friend Kevin's surprise party and drives home, but his reflexes are so impaired that he runs over a pedestrian on the way.²⁵ Fischer and Tognazzini write:

Whatever sort of control is required for moral responsibility, suppose that he didn't have that control at the time he ran over the pedestrian. In a straightforward sense, then, he couldn't help it. But is he therefore not responsible? Of course not, because in another straightforward sense, he *could* help it. After all, presumably he had control of whether to get drunk at the party in the first place. In this case, Dan is morally responsible for running over the pedestrian even though he [did not have whatever control is needed] at the time of the accident. To account for this, our theory of moral responsibility should include some notion of *tracing*. In this sort of situation, Dan's moral responsibility can be *traced back* to an action of which he had control—namely, his getting drunk. (author's italics).²⁶

This passage illustrates the concept of tracing, and we can see it is not a concept outside of normal standards. It is not a theoretically loaded idea and by that I mean compatibilism and incompatibilism can both make use of it. The only difference will be tracing an action back to processes which are deterministic or indeterministic and each side arguing for their conclusion. Also, we needn't trace every single action back to something prior as tracing only concerns those actions which are apt for moral assessment in terms of moral desert. If we can trace an action back to something in me that I control like a decision or choice then I am the source of the action. Of course, the decision or choice will be an undetermined one on a sourcehood incompatibilist view but sourcehood compatibilism will have tracing work in just the same way as a sourcehood incompatibilist view. We could say that Dan's decision is traceable back to his decision to get drunk which is presumably something that by 'normal' standards he controlled.

discussion than I offer see, Manual Vargas, "The Trouble with Tracing," *Midwest Studies in Philosophy* 29, no. 1 (2005): 269–91.; Seth Shabo, "More Trouble With Tracing," *Erkenntnis* 80, no.5 (November 2014): 987-1011.; Angela M. Smith, "Attitudes, Tracing, and Control," *Journal of Applied Philosophy* 32, no. 2 (2015): 115–32.

²⁵ Fischer and Tognazzini, "The Truth about Tracing,".

²⁶ Fischer and Tognazzini, "The Truth about Tracing," 532.

Dan is deserving of blame even if in the present moment he had no control because of a prior decision which he did control.

Leeway incompatibilism believes determinism is incompatible with the ability to do otherwise, since determinism entails there is only one possible future and thus only one possible course of action. As alternative possibilities are necessary for this view, agents lack free will and moral responsibility. As we will see in the next section libertarians are divided over which path to take. My own libertarian theory adopts a hybrid of these approaches but believes sourcehood is the more fundamental criterion. I offer a defence of sourcehood incompatibilism in chapter three.

A second key feature of the libertarian outlook is the denial of the equating of causation with necessitation. This denial of necessity should not be taken to be a denial of causation itself. While some libertarians deny that free acts can be caused or argue that they need not be caused, the review of literature I provide in this section shows this is just one possible approach. It is a mischaracterization to ascribe a belief in contra-causality to all libertarians.

Libertarians who do require actions and choices to be caused are committed to *probabilistic causation*. This is in opposition to *necessitarian causation*, the belief that an effect is necessitated by its causes. Libertarianism requires that there be some *indeterminism* in the causal history of action but indeterminism is not pure randomness. Elizabeth Anscombe defines indeterminism as the thesis, “that not all physical effects are necessitated by their causes.”²⁷

Anscombe does not deny some causes necessitate their effects, only that all causation is

²⁷ Elizabeth Anscombe, “Causality and Determinism,” in *Agency and Responsibility: Essays on the Metaphysics of Freedom*, ed. Laura Wadell Ekstrom (Boulder: Westview Press, 2001), 70.

necessitating. She has an example of a Geiger counter causing a bomb to explode if it registers a certain reading; whether it will or not is not determined, for it is so placed near some radioactive material that it may or may not register that reading. If the bomb does go off there is no doubt of the cause.²⁸

There are other examples we can use to that illuminate this point. Kane constructs a case of an angry husband who, while arguing with his wife, swings his arm down on her favourite glass table-top in an effort to break it. Kane postulates hypothetical indeterminism in the swing of the angry husband's arm so that it is literally undetermined whether the table will break up to the moment when it is struck. Whether the husband breaks the table or not is undetermined; and yet he is clearly responsible if he does break it. It would be a terrible excuse for the husband to claim, 'Chance did it, not me'. The presence of the indeterminism makes it possible that the husband could fail in breaking the table but since he did break the table, he is responsible, not chance.²⁹

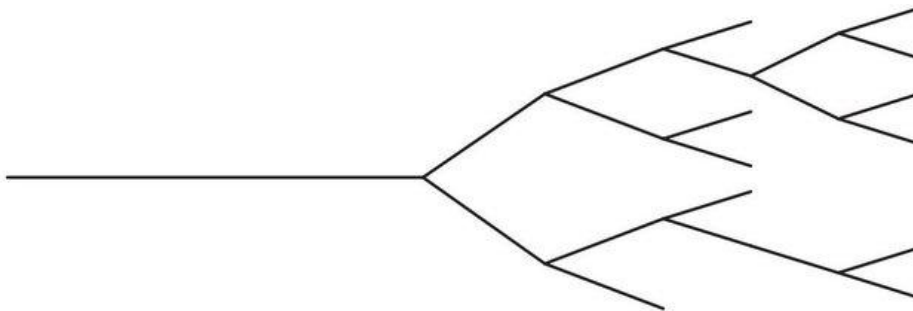
It needs to be noted that the exact role as well as the location of indeterminism is an internal dispute among libertarians.³⁰ A non-contentious conception of indeterminism that is broad enough to accommodate this fact is the idea that free agents face an *open future*. What is implied by the idea of an open future is illustrated by a metaphor, introduced into the literature

²⁸ Anscombe, "Causality and Determinism," 70.

²⁹ Robert Kane, "Acting 'of One's Own Free Will': Modern Reflections on an Ancient Philosophical Problem," *Proceedings of the Aristotelian Society* 114 (2014): 9.

³⁰ Christopher Evan Franklin, "How should libertarians conceive of the location and the role of indeterminism," *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* 16, no. 1, (Nov 2012): 445-8. ; Alfred Mele, "Two Libertarian Theories: Or Why Event-Causal Libertarians Should Prefer My Daring Libertarian View to Robert Kane's View," *Royal Institute of Philosophy Supplement* 80, (May 2017): 49-68.; Laura Ekstrom, "Indeterminist Free Action," In *Agency and Responsibility: Essays on the Metaphysics of Freedom*, ed Laura W. Ekstrom, 138-157. (Boulder: Westview Press, 2001): 138-157.

by John Martin Fischer, a semi-compatibilist, of a ‘garden of forking paths’, as seen in figure 1.1 below:³¹



As we can see, this picture of reality conceives that there are forks in the road that represent all the different choices we could make and the way the world could have been, and which possible world becomes the actual world depends on the choices we make. If determinism is true, however, the alternatives are only epistemic not metaphysical and there is only one possible future. As I shall demonstrate, all libertarians implicitly believe in a metaphysical picture like the garden of forking pathways. There is for libertarians something objectionable about one's life not having an open future in the sense just defined. An open future is important for libertarians because of its connection to the earlier point I made about an action being, “up to me.” It is in virtue of having a metaphysically open future that an action is “up to me.” The figure fits an understanding of “up to me” in terms of alternative possibilities but some libertarians believe an action being up to me articulates an open future in terms of ultimate sourcehood.

³¹ Robert Kane, “Libertarianism” in *Four Views on Free Will*. (Malden: Blackwell Publication, 2007), 167. Fischer concedes that there is something attractive to this picture of a metaphorical garden of forking pathways, and that the freedom to do otherwise is incompatible with determinism. He says, “It seems to me that our intuitive “picture” of the structure of possibility over time corresponds to Borges’ idea of the future as a “garden of forking paths.” That is, our intuitive conception or picture corresponds to a branching, treelike structure in which the various possible futures branch off a single line that can be traced back into the single actual past.” “Compatibilism” in *Four View on Free Will* (Malden: Blackwell Publication, 2007), 76. Fischer is a semi-compatibilist because he does find this picture of the freedom to do otherwise at odds with determinism, but maintains moral responsibility is compatible with determinism hence his semi-compatibilism.

3. Kanean Libertarianism

Kane charges libertarians with a task to articulate an intelligible and plausible incompatibilist libertarian theory. Kane describes what it will take for libertarians to accomplish this task with a metaphor he calls “Incompatibilist Mountain.” Kane says that a believer in libertarian free will is like a mountain climber in the sense that they have two mountain-climbing tasks. A libertarian must “ascend the mountain” which entails that they must demonstrate that free will is incompatible with determinism and that an incompatibilist free will is a significant kind of freedom. The second task libertarians have is to “descend” the mountain. Descending the mountain requires showing that an incompatibilist understanding of free will is intelligible and that such a freedom could exist and where to look for it. Kane describes the challenge for libertarians like this,

Imagine that the task for libertarians in solving this dilemma is to ascend to the top of a mountain and get down the other side.... Getting to the top of this mountain—demonstrating that free will and determinism are incompatible—is a difficult task for libertarians... But many critics of libertarianism believe the Descent Problem—making sense of a free will that requires indeterminism—is even more difficult and dangerous than the ascent; and this may be the case for libertarians. The air is thin and cold up there on Incompatibilist Mountain; and if you stay up for any length of time, say critics of libertarianism, your mind gets foggy. You start having visions of fantastical ideas, such as transempirical power centers, noumenal selves, and unmoved movers, which libertarians have often invoked to explain their view.³²

Kane begins his ascent up ‘Incompatibilist Mountain’ through disentangling the notion of freedom. This is where Kane introduces an important distinction between *free will* and *free action*. Kane introduces three ways in which we can speak of freedom regarding action and will:

(1) acts done voluntarily, on purpose and for reasons that are not coerced, compelled

³² Robert Kane, *A Contemporary Introduction to Free Will* (New York: Oxford University Press, 2005), 34.

or otherwise constrained or subject to control by other agents.

(2) acts [free in sense 1 that are also] done “of our own free will” in the sense of a will that we are ultimately responsible (UR) to some degree for forming.

(3) “self-forming” acts (SFAs) or “will-setting” acts by which we form and reform the will from which we act in sense 2.^{33, 34}

Kane defines *freedom of action* in terms of the first defined sense of freedom. To act freely is to be unhindered in the pursuit of your purposes. This kind of freedom is compatible with determinism. The free acts defined in two and three are exemplary of *freedom of will* in an incompatibilist libertarian sense because they could not exist in a determined world.

Kane states that libertarians should concede that free action in the first sense defined is a valuable form of freedom even if it is compatible with determinism. Kane believes what libertarians ought to do is to claim there is at least one kind of freedom that is incompatible with determinism, and that it is a “*significant kind of freedom worth wanting*.” (emphasis original)³⁵ For Kane, this additional incompatibilist freedom is “free will”, and it is the freedom to be “an ultimate creator and sustainer of one’s own ends or purposes. Those who defend such a freedom (“free willists,” as I call them) believe that free will in this sense is something worth wanting *over and above* compatibilist freedoms, such as freedom from coercion or compulsion or political oppression ...”³⁶ Free will as the power to be the “ultimate creator and sustainer of one’s

³³ Robert Kane, “The Complex Tapestry of Free Will: Striving Will, Indeterminism and Volitional Streams,” *Synthese* 196, no. 1 (2019): 147.

³⁴ Harry Frankfurt and Andrew Sneddon have discussed how coercion is compatible with voluntary action. See Frankfurt, Harry G, “Freedom of the Will and the Concept of a Person.” In *The Importance of What We Care About: Philosophical Essays*, (1988) 11-25. Andrew Sneddon, “Alternative Motivation and Lies” *Analysis* 81, no 1 (January 2021) 46-52.

³⁵ Kane, *The Significance of Free Will*, 14-15.

³⁶ Kane, *The Significance of Free Will*, 15.

own ends or purposes” is what Kane intends to express in the definitions of freedom in sense two and three. It is through SFAs in sense three that one becomes ultimately responsible and free in the second sense of the schema we are discussing. As I will argue in chapter five, I interpret this as having authorship over our lives and for the kind of persons that we are.

Kane’s view is a *source incompatibilist* one as he believes that free will, as well as moral responsibility, requires that persons are the ultimate source of their actions and, more fundamental, of the kind of person that they are. Only if we are ultimate sources of our identities can we be what Kane calls *ultimately responsible* for who and what we are.³⁷ Ultimate Responsibility (hereafter UR) is a crucial idea in Kane’s theory as he uses it to distinguish compatibilist freedoms from incompatibilist ones in the three previously defined types of freedom. I will have much more to say about UR and how to make sense of it in chapter four. As Kane conceives of it, to have UR is to have the freedom to choose to be the kind of person that one is, and he believes this is incompatible with determinism. For if determinism is true, we can never do otherwise throughout our entire life history, and this would mean we could never choose to be a different kind of person. We would therefore fail to be ultimately responsible for what and who we are.

The heart of Kane’s model is the idea of self-formation. Persons freely shape their wills through the performance of what Kane calls self-forming actions or SFAs. SFAs are how Kane believes agents become ultimately responsible. Kane’s view of the frequency of SFAs and when they occur is ambiguous. He says that SFAs occur at those difficult times in our lives when we are torn over what to do and who to become. Kane sometimes says that these identity implicating

³⁷ Kane, *The Significance of Free Will*, 36.

scenarios increase as persons get older because their lives increase in complexity with age.³⁸ A key feature of self-formation is the ambivalence in the minds of agents. Now how are we to understand this ambivalence? In the present view, the kind of ambivalence is not understood as mere indecision. The ambivalence has to be rooted in competing motivations internal to an agent's will. The reason we need to adopt this sense of ambivalence is that we need to be able to say the choice an agent makes is done in accordance with their wills and not against it.

Actions which are self-forming happen in ordinary circumstances and a fair question to ask is how frequently are they made and what is their nature? Up to this point they seem to involve tricky decisions but does this mean people who are indecisive make SFAs more than people who are more confident? SFAs are indeed tricky situations but not every tricky decision is necessarily an SFA. SFAs are decisions where one is not just torn in a sense of having to try really hard to make a decision. It runs deeper than that; one has an identification that comes from motivations which are internal to who one is at that time. So, if one is torn between two cars, that does not automatically make it an SFA. We can modify cases involving ordinary circumstances to generate a potential SFA but what is needed is a divide within oneself between personal motivations that are part of oneself.³⁹

We can illustrate SFAs with two examples. A businesswoman is walking to an important business meeting, and she witnesses an assault in an alley. An engineer is a recovering alcoholic trying to fix his marriage. He is working late and is under a lot of stress and the temptation to drink to relieve the pressure arises. They both face an inner conflict over what to do as each of

³⁸ Kane, "The Problem of Free Will: A Libertarian Perspective," 48.

³⁹ Kane, "The Problem of Free Will: A Libertarian Perspective," 48.

the options in the respective scenarios is attractive to the agents. The businesswoman faces a moral conflict, the engineer a prudential one. They, “are torn by conflicting motives and must make mental efforts to get their ends or purposes sorted out—to ‘set’ their wills in one way or another.”⁴⁰ Kane describes the cases respectively as the businesswoman being tempted to prioritize personal ambition over helping someone in distress and the engineer being tempted to relieve the present tensions.⁴¹ They are also, however, committed to their moral beliefs and long-term goals, and make mental efforts to not succumb to the temptations that threaten these commitments.

To stay with the businesswoman, she is torn over what to do. She sees that there are only two options: stay and help or continue on to her meeting. Such a person might deliberate over the conflicting motives and these motives are expressive of different aspects of her identity. She thinks to herself “do I want to ignore this person in trouble and be the kind of person who ignores others when their needs conflict with my own personal ambitions?” The businesswoman is motivated to help and to continue on to her meeting, and there are considerations which she is sensitive to that support choosing both of the options. So it’s plausible to say she chooses between moral reasons of helping someone in peril and to get to her meeting on time. Or alternatively we might describe her deliberation not as explicitly choosing to be a certain person but implicitly she is shaping her will through the performance of the action chosen. The businesswoman performs a “will-setting” action. An action is “will-setting” when an agent’s will is not already set one way or another.⁴² Self-forming action or SFAs are will-setting in this sense

⁴⁰ Kane, *The Significance of Free Will*, 126.

⁴¹ Kane, *The Significance of Free Will*, 126.

⁴² Kane, “The Problem of Free Will: A Libertarian Perspective,” 18-19.

because agents performing SFAs do not have their wills already set before they act. The businesswoman doesn't already have her mind made up or inclined one way before she acted. She can act intentionally and rationally either way and thereby set her will.

This seems like the appropriate place to introduce the nature and role indeterminism in Kane's account. On Kane's account indeterminism is described as an obstacle which must be overcome by an agent.⁴³ An agent like the businesswoman is faced with two options and because she is attracted to both alternatives, her pursuit of one goal is interfered with by her desire for the other goal. Indeterminism is not per se an obstacle, but rather it makes the situation described as a genuinely possible one. The businesswoman is tempted to go to her meeting and, she is tempted to stay and help, and the temptation to do one of these is an obstacle and indeterminism makes the obstacle more than psychological but genuine to her situation.

If the businesswoman overcomes the temptation to go to her meeting and stays and helps instead it will be the result of her mental effort. The point that needs emphasizing is that whatever the businesswoman does (and the engineer) they make one set of competing reasons prevail whichever ones they happen to be. Kane writes, "when agents, like the (business)woman, decide in such circumstances, and the indeterminate efforts they are making become determinate choices, they *make* one set of competing reasons or motives prevail over the others then and there *by deciding*." (author italics)⁴⁴ There is a sense in which failure is also involved in that if the businesswoman's reasons to stay prevail her reasons to go fail. By acting on one set of reasons

⁴³ Kane, *The Significance of Free Will*.; Kane, "The Problem of Free Will: A Libertarian Perspective."; Robert Kane, "On the Role of Indeterminism in Libertarian Free Will," *Philosophical Explorations* 19, no. 1 (2016): 2–16.

⁴⁴ Kane, "The Complex Tapestry of Free Will: Striving Will, Indeterminism and Volitional Streams," 150-151.

rather than another the alternative set of reasons fail. The reason Kane frames the model of deliberation this way is to emphasize that the agent is doing something and actively involved no matter what they choose. They make one set of reasons succeed; by deciding for those reasons they simultaneously make the other set of reasons fail.

It must be pointed out that “effort” for Kane, at least in cases of SFAs, is a technical term that Kane ties to a particular notion of will. There are different notions of will in his view and when we speak of what an agent makes an effort to do we are talking about a technical notion of an, “effort of will.”⁴⁵ Here are all the different meanings of will Kane uses:

- (i) what I *want, desire, or prefer* to do;
- (ii) what I *choose, desire, or intend* to do;
- (iii) what I *try, endeavor or make an effort* to do;⁴⁶

Kane writes that the first sense of will is *appetitive will* – willing-as-wanting-or-desiring-or-being-inclined to this or that results. The second meaning of will is what Kane calls *rational will*, will-as-practical reason, terminating in what agents choose, decide or intend to do. The third sense of will is what Brian O’Shaughnessy calls *striving will*. It is willing-as-trying-or-making-efforts.⁴⁷ Kane’s concept of “efforts of will” is an instance of striving will because it is an agent making an effort towards some goal. In the case of SFAs agents are making plural efforts of will. The businesswoman is making an ‘effort of will’ to get to her meeting on time and she is making an effort to help the person being assaulted. She makes these dual efforts because of the inner conflict between different aspects of her personal identity. She is a very ambitious career-driven

⁴⁵ Kane, *The Significance of Free Will*, 26.

⁴⁶ Kane, *The Significance of Free Will*, 26.

⁴⁷ Kane, *The Significance of Free Will*, 27.; Brian O’Shaughnessy, *The Will: A Dual Aspect Theory* (Cambridge: Cambridge University Press, 1980).

person but she is not unresponsive to the needs of others around her who need help. In the moment she has to make a choice on how to act. She has to make an effort to make one choice or the other. Whichever way an agent ends up acting it will be what they were trying to do as described by the third sense of will. This third sense of will is tied to SFAs because agents are making efforts or trying to bring about ends that are generated from their wills. The businesswoman wants to help and so makes the effort to decide to help but at the same time has to make the effort to resist the temptation to go on to her meeting. These conflicting efforts are expressive of multiple identities held together within one person with each one trying to overcome the opposing efforts made by others to make a choice.⁴⁸ In making an effort to choose the agent has to also make the effort to resist the other choice that they are making an effort to choose. The businesswoman is making conflicting efforts because part of her is making the effort to resist the other effort the other part of her will is trying to make. She recognizes genuine moral reasons to stay and to leave and recognizes the moral dilemma she has walked into. She thinks that it would be a bad thing for her to ignore the mugging victim and to be the kind of person that puts herself above everyone else; however, she also feels tempted to ignore someone that is a stranger to her so that she can get to her meeting. She explicitly feels torn about who she wants to be.

Now I think it is rare in actual cases to be so explicit in deciding what person to become. It is much more implicit, so how do we say I choose to become a kind of person without that

⁴⁸ Kane has changed his model regarding efforts of will. In his paper, "Making Sense of a Free Will that is Incompatible with Determinism: A Fourth Way Forward" Kane puts forth a new approach to efforts of will. Kane says, "it is not being claimed here (as I have done in earlier writings) that these efforts or exercises of will-power aimed at different choices would be occurring at the same time during deliberation. Nor will they be occurring throughout the entire deliberation. Rather, different efforts or exertions of will-power may be initiated at different times, depending on the course of the agent's reasoning or thought processes." As Kane indicates in the cited passage, in his earlier writing he did claim that the efforts of will involved in self-forming choices are efforts to do both A and B. I am describing Kane's view as he outlines it in *The Significance of Free Will*.

direct thought? We don't necessarily choose our identities directly but instead we form our identities through decisions and actions which shape who we are and become. The relationship between identity and choice is consequential as by making certain choices the effect would be to mold our identities. For instance, if Henry has taken on a diet regime with the goal to lose weight and he faces the temptation to break with his diet, he faces a prudential dilemma between sticking to his long-term goal and satisfying short-term pleasure. This is just one choice so his identity will not change much if at all but if we imagine many such temptations the consequence of always giving in to temptation shapes his identity. He shapes who he is consequentially by making himself inclined to give into temptation.

I thus understand Kane's view of free will to be deeply connected to personal identity. It seems to me that the essence of Kanean libertarian free will is that it requires that one create the kind of person you are and to be able to decide the kind of person you want to be. Moreover, the libertarian model Kane offers supports a view of persons as fluid in the sense that who they are is not fixed but an ongoing project formed and created through their free choices. Our identities are shaped in many ways but in an important sense they are shaped by the decisions made in SFAs. SFAs are choices that are not determined and yet an interesting consequence of them is that these are choices which determine the kind of person one is. This is why Kane can allow that alternative possibilities not be required for every choice to count as a choice made of our own free will. Through determining who we are through SFAs, choices and decisions which are not SFAs can consistently be said to not have alternative possibilities and be free.⁴⁹ What I mean here is that our free will can be preserved even when we cannot have done otherwise because we

⁴⁹ It is worth noting that on my view of personal identity passive factors can be part of our identities. I take a broad notion of personal identity where in addition to passive factors, we have an active components to our identities which is where SFAs come to be relevant. By having a notion of personal identity that has passive and active sense, I can consistently say there are parts which might be determined.

will have shaped the will from which we acted through performing SFAs. It should be emphasized that Kane's view does not entail that people determine who they are in one single SFA. The businesswoman does not entirely determine her character through a single SFA. She neither becomes completely indifferent to other people if she goes to her meeting nor does she stop caring about her career if she stays to help. It is possible that some agents might perform a transformative SFA but it is not necessarily true of every SFA and agent.

Kane's theory has faced many objections over the last four decades since he first conceived it.⁵⁰ The objection introduced in chapter one, "the luck objection" is the one in my view that is the most pressing.⁵¹ My solution to the 'luck objection' will come in chapter 6. In this chapter I want to focus on objections to Kane's view from other incompatibilist libertarian theorists. I propose a comparative method for assessing the alternative libertarian models in the literature. This will provide a rigorous picture of the options libertarians have crafted. In what follows I will construct the other three major libertarian models that have been developed and the challenges they pose to Kane's view. We will see the common threads throughout all these theories and what makes them libertarian. The differences will also come through and I believe that Kane's view will be shown to be the strongest one.

⁵⁰ Randolph, Clarke, *Libertarian accounts of free will* (Oxford: Oxford University Press, 2003); Fischer, Pereboom, and Vargas 2007; Franklin 2013; Carl Ginet, "Can an indeterministic cause leave a choice up to the agent?" *Libertarian Free Will: Contemporary debates*. Edited by David Palmer (Oxford: Oxford University Press, 2014), pp. 15-26; Meghan Griffith, "Why agent-caused actions are not lucky." *American Philosophical Quarterly*, 47 (January 2010): 43-56; Michael Mckenna, "Ultimacy and sweet Jane." in *Essays on free will and moral responsibility*. eds N. Trakakis & D. Cohen (Newcastle upon Tyne: Cambridge Scholars Publishing, 2008); Alfred Mele, *Free Will and Luck*. (Oxford: Oxford University Press, 2006); Aldina Roskies, "Can neuroscience resolve issues about free will?" in *Moral psychology* ed. Walter Sinnott-Armstrong (Cambridge, MA: MIT Press, 2014), 4, 103-126; Galen Strawson, "The Impossibility of Moral Responsibility" *Philosophical Studies*, 75, (August 1994), 5-24.

⁵¹ Mele, *Free Will and Luck*.; Mele, "Ultimate Responsibility and Dumb Luck" *Social Philosophy*, 16(2), (1999): 274-293; Mele "Kane, Luck, and Control: Trying to Get by without Too Much Effort", *Libertarian Free Will* (2014).

4. Alternative Libertarian Theories

I will focus on three major challengers to Kane's libertarianism. These are agent-causal, non-causal, and alternative event-causal theories.⁵² I will begin with analysing non-causal accounts before moving on to agent-causal theories. I claim these models are not preferable to the Kanean framework I've outlined. I'll begin with a brief overview of the basic tenets each competing theory holds before detailing the challenges levied against Kane by the proponents of these alternative theories.

a) *Non-Causal Libertarianism*

Libertarianism is sometimes thought to require that our actions must be uncaused in order to attribute them as free and morally responsible. This is untrue, however, as the majority of libertarians believe that free actions are caused by agents, albeit they disagree in what this entails. There is one group of hold outs, and they are non-causal libertarians. As the name might suggest, they do not merely make the familiar incompatibilist claim that free will is incompatible with determinism but go further and require free actions can *never be caused*.⁵³

Two of the leading non-causal libertarians, Carl Ginet and Hugh McCann, claim that what differentiates mere occurrences from actions are intrinsic properties actions possess that

⁵² David Palmer, "Introduction" in *Libertarian Free Will*. For those interested in surveying the full breath of the field a good place to start is "Libertarian Perspective on Free Agency and Free Will" Section IV in *The Oxford Handbook of Free Will* 2nd Eds Robert Kane (Oxford: Oxford University Press, 2012), 309-406 One kind of theory not found here is Christian List's "Libertarian-Compatibilist" view. His theory will also not be able to be addressed here but for readers interested to learn more see his *Why Free Will Is Real* (Oxford: Oxford University Press, 2019).

⁵³ Carl Ginet, "Can an indeterministic cause leave a choice up to the agent?" in David Palmer eds, *Libertarian Free Will: Contemporary Debates*. (New York: Oxford University Press, 2014); Carl Ginet *On Action* (Cambridge: Cambridge University Press, 1990); Stewart Goetz, *Freedom, teleology, and evil*. (London: Continuum, 2008); Stewart Goetz "A noncausal theory of agency" *Philosophy and Phenomenological Research*, 49, no. 2 (Dec,1998) 303-316; Stewart Goetz "Libertarian choice." *Faith and Philosophy*, 14, no. 2 (April 1997), 195-211; Hugh McCann, "Making decisions" *Philosophical Issues*, 22, no.1 (October 2012), 246-263; Hugh McCann, *The Works of Agency: On Human Action, Will, and Freedom* (Ithaca: Cornell University Press, 1998)

mere occurrences do not.⁵⁴ For instance Ginet believes that actions have an intrinsically “actish phenomenal quality” which agents feel when acting.⁵⁵ The following passage, while lengthy, helpfully discusses Ginet’s view of action and freedom and shows what he requires for both.

Ginet writes,

Every action, according to me, either is or begins with a causally simple mental action, that is, a mental event that does not consist of one mental event causing others. A simple mental event is an action if and only if it has a certain intrinsic phenomenal quality, which I've dubbed the "actish" quality and tried to describe by using agent-causation talk radically qualified by "as if": the simple mental event of my volition to exert force with a part of my body phenomenally seems to me to be intrinsically an event that does not just happen to me, that does not occur unbidden, but it is, rather, as if I make it occur, as if I determine that it will happen just when and as it does (likewise for simple mental acts that are not volitions, such as my mentally saying "Shucks!"). A simple mental event's having this intrinsic actish phenomenal quality is sufficient for its being an action. But its having the quality entails nothing either way as to whether it satisfies the incompatibilist requirement for free action (which is that it not be causally necessitated by antecedent events). An action may be causally complex, may consist of a simple mental action plus consequences of it. For example, my action of voluntarily pushing with my arm and hand against a door begins with a volition, a simple mental act of willing to exert a certain force in a certain direction with my arm and hand, and consists further in that volition's causing my arm and hand to exert such a force. My action of opening the door has a still further component of the door's opening being caused by the force exerted against it by my arm and hand. Now, as I explained earlier, if an event is not an action of mine — for example, the door's opening — then I can make that event occur only by causing it, that is, by performing some action that causes it. But I make my own free, simple mental acts occur, not by causing them, but simply by being their subject, by their being my acts. They are ipso facto determined or controlled by me, provided they are free, that is, not determined by something else, not causally necessitated by antecedent states and events.⁵⁶

The perennial objection pressed against this kind of libertarianism is the “control objection”. David Palmer puts the worry in the following way:

⁵⁴ Ginet, *On Action.*; McCann, “Making Decisions.”

⁵⁵ Ginet, *On Action*, 11-15.

⁵⁶ Ginet “Freedom, Responsibility, and Agency” *The Journal of Ethics* 1 no. 1 (September 1997): 89.

If a person's actions *were* uncaused (as the noncausal libertarian view requires), how can she have the required control over whether they occur in order for her actions to be free?⁵⁷

Free actions must be under our control in the sense that we at least control the initial spring of action. If an action is uncaused this would seem to entail that an agent has no control over an action they perform because actions would just occur. This would entail no one could control whether an uncaused action occurs or not occurs. Therefore, a free action *cannot* be uncaused, contrary to noncausal libertarian's assertions.⁵⁸ I find this objection very powerful and decisive against the non-causal libertarian. Human beings are part of the fabric of the natural world and therefore we are governed by causality as much as anything else. Why should we want a noncausal freedom? Drawing out Ginet's account can help us to get a better sense of what noncausalists value about noncausal freedom.

Ginet's argument is similar to Peter Van Inwagen's Consequence Argument, which claims that if determinism is true, then everything in the present is a consequence of the past combined with the laws of nature.⁵⁹ Ginet spells out his argument in similar terms to the Consequence Argument with the use of logical principles that he believes entail incompatibilism.⁶⁰ Ginet's argument for the incompatibility of free will and determinism draws on the previously discussed notion of free actions being 'up-to-me'. Ginet seems to understand an action being 'up to me' in the leeway sense. He argues that if determinism is true then an agent does not have it open to them to either make it the case that *p* or pursue alternative courses

⁵⁷ David Palmer, "Free will and control: A noncausal approach" *Synthese* 198, no. 10 (2021): 10043–62.

⁵⁸ Palmer, "Free will and control," 3.; See also, Randy Clarke, *Libertarian Accounts of Free Will*, (New York: Oxford University Press, 2003).

⁵⁹ Peter Van Inwagen, *An Essay on Free Will* (Oxford: Clarendon Press, 1984).

⁶⁰ Ginet, *On Action*, 102.

of action.⁶¹ Thus, Ginet's leeway incompatibilism cashes out its understanding of 'up-to-me' as the truth of an event being contingent on whatever course of action I take. So, if an action must be 'up-to-me' then what makes it the case that p must be *me*. Consider the following case involving Jones and Smith where Jones kills Smith. If it was open to Jones to kill Smith it was open to him to make it the case that p , that he would kill Smith. He could have refrained from killing Smith making it 'up-to-him' whether he would indeed kill Smith. Ginet's argument for incompatibilism would state that if determinism is true, it follows that it was already the case that p , (that Jones would kill Smith) before Jones ever even thought about it.⁶² So it was already the case that p , it was a fact Jones would kill Smith. Since no one has any control over whether something is a fact it is unavoidable that Jones will kill Smith.

This argument is only meant to establish incompatibilism and it alone will not establish libertarian free will. The further condition needed on Ginet's view is unsurprisingly that an action be uncaused. This can be understood by extending the above argument in the following way: an agent freely performs an action if and only if that the agent is the reason that p is the case and p is not made to be the case by any antecedent causes. Only if an action is uncaused can it be established that it was Jones who made it the case that p . As Palmer notes, Ginet, as well as other noncausalists, only posit negative requirements for an action to be free.⁶³ Ginet's view states what cannot be true for an agent to exercise free will but this does not get us very far, if

⁶¹ Ginet, *On Action*, 102.

⁶² Ginet, *On Action*, 98-100.

⁶³ Palmer, "Free will and control," 3.

anywhere. What exactly does a noncaused action accomplish in securing free will or moral responsibility?

Interestingly, Ginet is concerned with both determinism and indeterminism. His strategy is one that attempts to reconcile free will with both determinism and indeterminism. His account of free agency and free will believes that our freedom is threatened by both deterministic and indeterministic causation. Offering his thoughts on the luck objection, Ginet writes,

If there is a genuinely appealing intuition behind the luck argument, it is this: if a decision is caused indeterministically, then what the agent decides is not determined by the agent at the time of the decision but by nature's "roll of the dice," where the outcome probabilities are determined by the antecedent state of the world and the laws of nature. But the basic intuition here, it seems to me, applies to the case of causation generally, whether deterministic or indeterministic—it is nature that decides what its current state will produce next, whether by a deterministic or by an indeterministic "roll of the dice" (a deterministic roll being one where all the faces of the dice have the same number of spots on them). But the basic intuition does not apply to the case where the decision is not caused at all. There nothing decides what the choice is but the choice itself—that is to say, the chooser in making the choice.⁶⁴

Ginet appears to feel the force of the intuition that "whether deterministic or indeterministic—it is nature that decides what its current state will produce next, whether by a deterministic or by an indeterministic 'roll of the dice' (a deterministic roll being one where all the faces of the dice have the same number of spots on them)." I think Ginet is mistaken to say indeterministic causation is the same as deterministic causation in undermining our freedom and moral responsibility.

Ginet's account would allow us to say there are mental events involved in free choices but that the production of choice is not caused by these events. So, if I choose to have pancakes for breakfast we can say I had a mental event like a desire for pancakes but on the non-causal

⁶⁴ Ginet, "Can an indeterministic cause leave a choice up to the agent?," 24.

account we cannot say that the event caused my choice. Suppose that my choice of breakfast was indeterministically caused by my mental event of coming to desire pancakes. Suppose further I faced a choice between pancakes and waffles and there is a corresponding mental event of desiring waffles. My choice of breakfast is undetermined so I can have pancakes or waffles since I am able to do otherwise and choose the alternative. That the explanation for my choice is in terms of indeterministic causation producing my choice does not seem to undermine the claim that I decided what to eat and not nature. It seems to me that the intuition Ginet appeals to contains an assumption about indeterminism, which is that indeterminism is incompatible with an agent being able to decide. Why? There seems to be two theoretical possibilities.

The first is that an agent does not decide what they do because they lack control over their decision because they are not able to choose to act or to choose to refrain from acting. Indeterministic causation on this line of thought seems to make the indeterministic cause the reason an agent chooses and not the agent herself. So, in order for an agent to decide, her decision cannot be caused by anything prior to her act of deciding. This seems vulnerable to the objections that philosophers have made about noncausal libertarianism which is cast in terms of control. R.E. Hobart argues that if actions were uncaused they could not be free because they would spontaneously occur with no input from the agent.⁶⁵ We could not control our actions if they just happen *ex nihilo*. Laura Ekstrom sharpens this criticism by saying that there would be no connection between agents and actions because actions would not be produced by anything at all.⁶⁶ David Velleman observes that on the noncausal libertarian view, agents might feel like they

⁶⁵ R.E. Hobart, "Free Will as Involving Determinism and Inconceivable Without It," *Mind*, 43, no.169 (January 1934): 1-27.

⁶⁶ Ekstrom, "Indeterminist Free Action," 140.

are making free choices but if free choices cannot be caused and are just produced out of nothing then they are just suffering an illusion that they are in control.⁶⁷

The second possibility seems to be that if (deterministic or indeterministic) causation is involved the agent is not able to decide because they are not the source of the decision to act. But why can an agent not be the source of her decision if it is indeterministically caused? If we locate the indeterministic causation in the decision making process so that what an agent decides is undetermined even though caused the agent will still play a role in producing her action in a causal way. Noncausal libertarianism can allow that my choice is an event and that I have mental events prior to choice but there is no causal relationship between the two; it is not apparent what advantage this gives noncausal views over causal ones in securing sourcehood. Am I not the source of my desires which on a causal view cause my actions? Ginet's view seems to suggest that every action begins with one mental event and nothing else prior, but this would mean that there are countless causal chains that come into being *ex nihilo*, which is at best implausible if not just straightforwardly false.

For the sake of argument let's say these criticisms can be solved. I want to turn back to my earlier question of what an uncaused event adds to libertarianism. Noncausal accounts like Ginet's appear to treat indeterminism as serious as a problem for free will as determinism so they create a view that shuns causation from freedom. In my view, it is a mistake to go the noncausal route. Libertarians should not try to solve the problem that indeterminism creates for their view by abandoning all forms of causality. Instead, we should attempt to make sense of how indeterminism works in the natural world. Since human agents are part of the same world, the

⁶⁷ David Velleman, "What Happens When Someone Acts?," *Mind* 101, no. 403 (1992): 461–81.

task for libertarians is to articulate how agents act freely in an indeterministic world.⁶⁸ Noncausal libertarianism attempts to do this, it seems to me, by dodging the problem altogether.

Libertarians must face the problem head on and not create more problems for themselves than necessary. Requiring mysterious noncaused events creates more puzzles and obscures rather than illuminates a path to an intelligible and reasonable libertarianism.

b) Agent-Causation

Timothy O'Connor, following supporters in the tradition like Thomas Reid, believes that agent-causation is, "the power of an agent to cause an intention in order to satisfy some desire or to achieve some aim."⁶⁹ A key concept in understanding this is the idea of *substance* causation. Agent-causal theories maintain that agents are substances with causal powers or properties. O'Connor believes that these properties are irreducibly dispositional. He describes them as tendencies to interact with other qualities in producing some effect, or a range of possible effects, and he allows they may be probabilistic but there are limiting determined cases.⁷⁰

O'Connor's variety of agent-causation is heavily linked to a powers-based view of causation. In his view, agent-causation is,

an ontologically primitive type of causation, one that is uniquely manifested by (some possible) persons and is inherently goal-directed and nondeterministic. It is not directed to any particular effects, but instead confers upon an agent a power to cause a certain type

⁶⁸ Assuming that the world is verified to be indeterministic as the standard interpretation of quantum mechanics says it is.

⁶⁹ Timothy O'Connor, "Agent-Causal Power." In *Free Will*. Second Edition. ed. Derk Pereboom (Indianapolis: Hackett Publishing, 2009): 308–328.

⁷⁰ Timothy O'Connor, "Agent-Causal Theories of Freedom." In *The Oxford Handbook of Free Will*. Second Edition. eds. Robert Kane, (New York: Oxford University Press, 2011): 309–328.

of event within the agent: the coming to be of a state of intention to carry out some act, thereby resolving a state of uncertainty about which action to undertake.⁷¹

O'Connor's account is heavy on metaphysics in the sense that his view of free will depends upon his account of a metaphysical model of causation being correct. Other free will theorists do not take this approach, as people like Kane focus on free will independent of the issues surrounding the metaphysics of causation.⁷² O'Connor strongly disagrees with an independent methodology as he believes any sufficient solution to the problem of free will must, "squarely engage these more fundamental metaphysical issues (the metaphysics of causation and the ontology of mental states)—they cannot be 'bracketed off' in the way that even some contemporary theorists of free will tend to suppose."⁷³

The fundamental pillar of this view is substance causation or, as Roderick Chisholm calls it, immanent causation.⁷⁴ The difference between event causation and agent-causation is the reduction of agents or substances to events. Event-causation casts explanations of actions into causal terms that relate the cause of an action to other events or states of affairs. Proponents of agent-causation resist this move and assert that it is agents or substances that cause an event or events responsible for action(s). Chisholm describes the two like so, "I shall say that when one event or state of affairs (or set of events or states of affairs) causes some other event or state of

⁷¹ O'Connor, "Agent-Causal Theories," 314.

⁷² As Kane himself says, his view is to be differentiated from theories like O'Connor's which Kane calls AC (nonoccurrent causation) theories as his view is a "TI" or teleological intelligibility theory. The difference being that the choices agents make are cast into a teleologically intelligible whole as opposed to casting the explanation in terms of a substance causing an event to occur. Kane, *Significance of Free Will*, 146.

⁷³ Timothy O'Connor, "Free Will and Metaphysics" in *Libertarian Free Will: Contemporary Debates*. First Edition. eds. David Palmer (New York: Oxford University Press, 2014): 27-36.

⁷⁴ Roderick Chisholm, "Human Freedom and the Self" in *Free Will* eds. Robert Kane (Malden: Blackwell Publishing, 2002): 52

affairs, then we have an instance of *transeunt* causation. And I shall say that when an *agent*, as distinguished from an event, causes an event or state of affairs, then we have an instance of *immanent* causation”⁷⁵ There are differences in the details of traditional agent-causal accounts in how to spell out agent-causation.⁷⁶ What is common to them all though is, “that when an agent acts with free will, the event that is directly agent-caused has no complete event-cause. Only if there is such an absence of event causation, it is said, can there be two or more courses of action that are genuinely open to the agent; only then can it be assured that when an agent acts, it is genuinely up to the agent what she does.”⁷⁷

An important exception to this is Helen Steward and her “Agency Incompatibilism.”⁷⁸ Steward is and is not an agent-causationist as it depends on how one defines it. If it is the claim that agents cause their actions or decisions and nothing else, then she definitely rejects it.⁷⁹ Her view is in a sense much stronger than a lot of other incompatibilists as she argues that agency itself, defined as a power or set of powers and attributed to both human and non-human animals, is incompatible with determinism.⁸⁰ Steward’s approach to the issue of agent vs event causation

⁷⁵ Chisholm, “Human Freedom and the Self,” 52.

⁷⁶ See Richard Taylor, *Action and Purpose* (Englewood Cliffs, N.J.; Prentice-Hall, 1966): 12; Roderick Chisholm, “Freedom and Action,” in Keith Lehrer, ed., *Freedom and Determinism* (New York: Random House, 1966); Alan Donagan, *Choice: The Essential Element in Human Action* (London: Routledge and Kegan Paul, 1987); Timothy O’Connor, *Persons and Causes*. (Oxford: Oxford University Press, 2001); Thomas Reid, *Essays on the Active Powers of the Human Mind*, in Sir William Hamilton, ed., *The Works of Thomas Reid, D.D.*, vol. 1, 3rd ed. (Edinburgh: Maclachlan and Stewart, 1852).

⁷⁷ Randolph Clarke, “Agent Causation and Event Causation in the Production of Free Action.” *Philosophical Topics*, 24, no. 2 (1996): 20.

⁷⁸ Helen Steward, *A Metaphysics for Freedom* (Oxford: Oxford University Press, 2012).

⁷⁹ Steward, *A Metaphysics for Freedom*, 199.

⁸⁰ Steward, *A Metaphysics for Freedom*, 1.

is to adopt causal pluralism.⁸¹ Other philosophers have also followed suit in opting for this kind of pluralism. It must be noted that causal pluralism understands different types of causation not in competition but as independent from one another.⁸² Steward essentially denies the dichotomy between having to choose event-causation and agent-causation as she believes causation is a broad *category*: a large umbrella under which, “we bring a wide diversity of ontologically various relations and relationships, unified only by their connections to our interest in the explanation, prediction, and control of phenomena.”⁸³

C.D. Broad’s often cited objection to the possibility of agent-causation is relevant here.

Broad writes,

The putting forth of an effort of a certain intensity, in a certain direction, at a certain moment, for a certain duration, is quite clearly an event or process, however unique and peculiar it may be in other respects. It is therefore subject to any conditions which self-evidently apply to every event, as such. Now, it is quite evident that, if the beginning of a certain process at a certain in time is determined at all, its total cause *must* contain as an essential factor another event or process which *enters into* the moment from which the determined event or process *issues*. I see no *prima facie* objection to there being events that are not completely determined. But, in so far as an event *is* determined, an essential factor in its total cause must be other *events*. How could an event possibly be determined to happen at a certain date if its total cause contained no factor to which the notion of date has any application? And how can the notion of date have any application to anything that is not an event?⁸⁴

Ekstrom levies pretty much the same objection against agent-causal views when she says,

⁸¹ Steward writes, “My own view, in fact, is that the question what the relata of the causal relation are... is simply not a sensible one.... There simply is no single relation that is *the* causal relation.” Steward, *A Metaphysics for Freedom*, 210.

⁸² Randolph, Clarke and Justin Capes, "Incompatibilist (Nondeterministic) Theories of Free Will", *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/fall2020/entries/incompatibilism-theories/> .

⁸³ Steward, *A Metaphysics for Freedom*, 210.

⁸⁴ Charlie Dubar Broad, “Determinism, indeterminism, and libertarianism.” In *Ethics and the History of Philosophy: Selected Essays*. (New York: Humanities Press, 1952), 215.

Take a case of a putatively free act, in virtue of having at its core an agent-caused event: say, my making a promise to myself not to eat so much chocolate. I, the agent, existed prior to the occurrence of this act, and I continue to exist after I have made the promise to myself. But if I exist both prior to the act's occurring and after it is done, without the act's occurring at any of those other times, then what makes the difference at the time it does occur? How can *I* as a persisting entity make something happen (or come to exist)? Normally when something happens, something else happened previous to it to cause it to happen... But I do not happen; I simply exist. And I exist both before and after the event in question. So what caused the promise to stop eating so much chocolate to occur when it did?⁸⁵

The objection to agent-causal views is that if we conceive of persons as substances with causal powers to irreducibly cause their actions then actions become mysterious. For if agents cause their actions through a brute relation between them and their acts the question raised is why did they act then rather than earlier or later? Ekstrom's version raises the important point that when conceived as substances agents seem to come in and out of causal chains as if they continuously create new events or states of affairs *ex nihilo*.

My own concern with agent-causal libertarianism is that I do not see what it adds to the libertarian position. For starters, we do not need to postulate a special causal relation in order to affirm that there are agents that act in the way that libertarians believe they do. What is it about agent-causation that without it we could not say that agents make choices, self-form their characters and motives and be ultimately responsible for who and what they are? Agent-causation is posited as a special relation, but it is not at all clear what it allows us to say that occurrent or event causation does not. I see no good reason to deny that agents pursue goals and direct themselves simply because we can say what it is about agents that make them pursue these goals in terms of events. If libertarians are to ever dispel the not completely unwarranted charge

⁸⁵ Laura Ekstrom, "Indeterminist Free Action" in *Agency and Responsibility: Essays on the Metaphysics of Freedom*, eds Laura Ekstrom (Boulder: Westview Press, 2001): 142.

of obscure or “panicky metaphysics”, to use P.F. Strawson’s phrase, I believe they must avoid taking on unnecessary ontological relations or entities.⁸⁶

c) *Event-Causal Libertarianism*

There have been various event-causal libertarian accounts developed in debates about free will and moral responsibility, with some proposed but not endorsed by non-libertarians.⁸⁷ I will deal with two prominent event-causal libertarian accounts from Laura Ekstrom and Christopher Evan Franklin respectively.⁸⁸ I am focusing on these two accounts as they are the direct competition to Kane’s libertarianism in the sense that they are all event-causal libertarians, and no other event-causal view is as developed as Ekstrom and Franklin’s views. I will begin with Franklin who has engaged in a debate with Kane over the role and location of indeterminism in libertarian theories of free will and moral responsibility. Through engaging in the debate between Franklin and Kane I aim to illuminate the relationship between libertarianism and indeterminism. Ekstrom’s view shares a high level of overlap with Kane and discussing her theory is instructive in showing why Kanean libertarianism is the best libertarianism on offer.⁸⁹

⁸⁶ Peter Frederick Strawson “Freedom and Resentment” in *Freedom and Resentment and Other Essays*. (London: Routledge, 2008.) I have in mind Cartesian egos or souls and Kantian noumenal selves.

⁸⁷ One example of a non-libertarian philosopher who develops a libertarian account without endorsing it is Alfred Mele, “Two Libertarian Theories: or Why Event-Causal Libertarians Should Prefer My Daring Libertarian View To Robert Kane’s.” *Royal Institute of Philosophy Supplement* 80, (May 2017): 49-80.

⁸⁸ Laura Ekstrom, “Toward a plausible event-causal indeterminist account of free will” 2016, *Synthese*, 196, 127-144; *Free Will: A Philosophical Study* (Boulder: Westview Press, 2000; Christopher Evan Franklin, *A Minimal Libertarianism: The Promise of Reduction*, (New York: Oxford University Press, 2018).

⁸⁹ I should note that I will not be discussing Mark Balaguer’s libertarian theory in *Free Will As An Open Scientific Problem* Cambridge, Mass: MIT Press, 2010 because while his view might be construed as an event-causal libertarian account he does not embrace the label himself. He believes that the compatibility question is an entirely a semantic one and is metaphysically irrelevant to the free will debate. All that matters, for Balaguer, is that individual human agents are what he calls “L-free.” Agents are L-free iff there is some indeterminacy in the causal processes that generate action in the appropriate way. These claims about the metaphysical relevance or irrelevance of compatibilism would take us too far off track and so I am not going to delve into his account. Moreover, a

d) A Minimal Libertarianism

Franklin's minimal libertarianism is a direct challenge to Kane's account. Franklin believes that Kane's basic recipe is mistaken and has fanned the flames of the luck argument.⁹⁰ His fundamental objection to Kane's conception of the role and location of indeterminism in his libertarian theory.

Franklin writes,

Over the past thirty years, Robert Kane ... has developed a highly influential event-causal libertarian model, which has, rightly, become the standard-bearer for event-causal libertarianism.... I contend that there are decisive reasons for preferring my minimal event-causal libertarianism to Kane's: minimal event-causal libertarianism is able to solve the problem of luck and the problem of enhanced control within a simpler and more intuitive model of free will.⁹¹

Franklin is sympathetic to yet critical of Kane's account because of his recipe for solving the problem of luck. Franklin argues that the recipe involves three parts:

- (1) Kane's claim that indeterminism is a hindrance to control;
- (2) Kane locates the indeterminism between efforts of will and the choice made;
- (3) The agent's contribution to their free choices is reducible to states of, and events involving, the agent.⁹²

Since an effort of will is itself a mental action, Kane thus locates the indeterminism between actions. Some further exposition of Kane is needed to understand Franklin's argument.

substantial part of Balauger's discussion is in answering what we saw Kane call the 'Existence Question' and I am not interested in providing an answer to that here.

⁹⁰ Christopher Evan Franklin, "How should libertarians conceive of the location and role of indeterminism?" 2012 *Philosophical Explorations*, 16, no. 1 (2012): 45; Franklin, *A Minimal Libertarianism*.

⁹¹ Franklin, *A Minimal Libertarianism*, 93.

⁹² Franklin, "How should libertarians conceive of the location and role of indeterminism," 48.

Recall the angry husband case from my earlier discussion of Kane. The husband is trying to break his wife's favourite table to spite her. Kane understands the presence of the indeterminism, "not as a cause *acting on its own*, but as an *ingredient* in larger *goal-directed* or *teleological* activities of the agent, in which the indeterminism functions as a *hindrance* or *interfering* element in the attainment of the goal. The choices that result would then be *achievements* brought about by the goal-directed activity (the effort) of the agent, which might have failed since it was undetermined, but did not"⁹³(authors italics). There is an important point being advanced here: first is that the husband's responsibility and control for breaking the table are not incompatible with his action being undetermined. Secondly, the nature of indeterminism postulated here is one where indeterministic processes that are involved in the production of certain choices involve agents in such a way that what they do is quite clearly the result of what they wanted and tried to do. The sense of an obstacle is not indeterminism per se but the effort needed to make success occur rather than failure. Indeterminism is just an ingredient and what it adds to the recipe is the chance of failure needed to be overcome. This is quite evident for the angry husband case. He might have failed to swing his arm with enough force to break the table because of the indeterminism present but he was successful despite this possibility and so it would be false to say he was not responsible. Indeterminism does not undermine our control or responsibility because it does not result in actions resulting in random outcomes or agents performing actions randomly. The processes that produce action which are indeterministic include the goals of agents directing the production of action. To put it a different way, the agent's goals themselves complicate the pursuit of each other. I am attracted to A and B and my

⁹³ Robert Kane, "On the Role of Indeterminism in libertarian free will," *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action* (2016): 5.

desire for B complicates my pursuit of A and indeterminism makes it open to me to genuinely choose A and B. This is a vital point for libertarian free will if we are to understand how indeterminism factors into free will without making a problematic criterion.

Another example Kane uses to bolster this point is a case, inspired by J.L. Austin, of an assassin trying to murder a prime minister.⁹⁴ Kane imagines that there is some indeterminism present in the assassin's nervous system in the form of a nervous twitch. Therefore, the assassin could fail in what he is trying to do, to murder the prime minister. If he succeeds in achieving his goal and overcoming the obstacle the indeterminism presents for him, he will be morally responsible for what he was trying to do. It would be a poor excuse indeed if he said because *he could have failed* in murdering the prime minister that he is not blameworthy *for* the murder. His will was set in one way so that he voluntarily, intentionally, and purposely tried to kill the prime minister; if he succeeds it will be exactly what he made an effort or tried to do.

Kane uses these cases to demonstrate that an agent can be morally responsible for undetermined actions. It seems false that either the husband or the assassin was not in control just because of the presence of indeterminism. That it was an obstacle they had to overcome to achieve their goals (as Kane believes) seems to reinforce this point because they had to struggle to bring about their desired ends. An agent who tries and succeeds to bring about a particular goal seems on the face of it, all else being equal, to be in control of and over their actions as well as deserving of moral praise or blame. Of course, there might be instances of weird luck in some cases, but then we would be in control and morally responsible for intending to bring something about even if the luck involved robs me of control of the outcome. Simply because an action is

⁹⁴ J.L. Austin's "Ifs and Can's. In *Philosophical Papers*, edited by J. O. Urmson and G. Warnock, (Oxford: Clarendon Press, 1961): 153-180.

undetermined does not entail an agent has no control or lacks moral responsibility for the action performed. If I act and no odd things happen even though they might, the fact that I tried to make it happen is enough, assuming all else is equal, to assign control and moral responsibility.

Here's another case from Kane that strengthens this point. Suppose there are three assassins each trying to kill a prime minister. Imagine each assassin has some indeterminism present inside them so that it is undetermined if they will succeed in hitting their target and therefore possible they might fail. Kane supposes each assassin has a different level of probability of success in achieving their goals. The first assassin has an 50% chance of success, the second has a 80% chance, and the third has a nearly 100% chance of success. If all three assassins succeed, they are all *equally* responsible. It would be absurd to give them 50 years, 80 years, and 100 years in jail because of the respective differences in the chance of success.⁹⁵ If they all succeeded, they were all doing what they wanted and were trying to do. It cannot be they lacked control or moral responsibility simply because it was undetermined whether they would succeed in accomplishing their goals. An interesting feature of Kane's strategy here is the way it parallels compatibilist arguments regarding determinism, responsibility and control. Suppose that it was determined that the husband would break his wife's table. Compatibilists could argue that the deterministic causal chains included the husband's decision to break the table and he has control for his actions and is blameworthy for them. The husband did exactly what he wanted to do in breaking the table; that he could not have done otherwise is irrelevant say compatibilists. But in the case where the outcome is undetermined the table breaking is up to the husband; this is the only difference between the two cases.

⁹⁵ Kane, "Libertarianism," in *Four Views on Free Will*. First Edition (Malden, Massachusetts: Blackwell Publishing, 2007): 35.

Franklin believes that this strategy is flawed because the husband and assassin cases assume that they are free and responsible for making the effort of will to make the choices they did. Franklin claims that this assumption is problematic because it places Kane in the following dilemma: on the first horn of the dilemma, cases like the businesswoman locate responsibility for action in responsibility for agent's efforts. But we can trace the efforts back to earlier SFAs and the efforts made to perform those earlier SFAs, and those efforts can be traced still further to even earlier SFAs, and we can trace efforts to choices which are not SFAs and which we could not control. The second horn of the dilemma bites the bullet and accepts that we do not control the efforts we make in SFAs, but this means that the way Kane constructs his model is problematic. Kane locates responsibility for actions in SFAs in the efforts of will agents make because even though the act they perform is undetermined they are responsible should they succeed in doing what they were trying to do. Indeterminism in Kane's view is conceptualized as an obstacle which he claims diminishes control in the previously discussed model of antecedent determining control where agents can guarantee that they perform or refrain from performing an action. This is seen by Franklin to be a problem because Kane appears to be conceding that indeterminism does diminish control and libertarianism is supposed to show how indeterminism enhances control. Through proposing a model of indeterminism that diminishes control, Kane's account fans worries about libertarian free will being undermined by luck. Saying indeterminism diminishes control seems to make undetermined actions reduce to luck.

Franklin begins setting out on reforming Kane's account by dropping or demoting Kane's efforts of will. Franklin agrees that agents make efforts to make up their minds over what to do

and to follow through on their decisions but denies that these are ‘efforts of will.’⁹⁶ He replaces Kanean efforts of will with his preferred concept of basic action.⁹⁷ Basic action is an action which does not require the performance of another action.⁹⁸ For example, turning on the light switch and startling a burglar. The movement of one’s finger is the basic action, and the turning on of the light and the startling of the burglar are increasingly less basic. This definition is for exposition purposes and I make no real commitments to the nature of basic actions. Free will is located at moments of undetermined basic actions. The indeterminism in basic action is understood not to be an obstacle to be overcome as it is in Kane’s account. Franklin articulates a conception of indeterminism where the efforts or choices we make are undetermined. Indeterminism opens up alternatives for agents but Franklin’s proposal does not concede that indeterminism diminishes control.

Franklin also abandons Kane’s idea of SFAs and choosing between incommensurable alternatives.⁹⁹ With all these pieces in place here is Franklin’s formal statement of his view,

Minimal event-causal libertarianism: An agent S ’s action ϕ at time t was directly free and one he was directly morally responsible for iff (i) S was normatively competent with respect to ϕ at t , (ii) ϕ was a basic action, (iii) S ’s reasons that favored ϕ nondeviantly brought about ϕ at t , and (iv) it was possible, given the past and laws of nature up until t , that R not have caused ϕ .¹⁰⁰

⁹⁶ Franklin, “How should libertarians conceive of the location and role of indeterminism?,” 52.

⁹⁷ Franklin, “How should libertarians conceive of the location and role of indeterminism?,” 52.

⁹⁸ Some theorists hold that there is a special class of actions out of which all other actions are composed, or which form the foundation for all actions. These are the basic actions. Typical candidates for basic actions are bodily movements. Other theorists contend that there is no such special class. Instead, whether an action is basic is a contingent issue. There are basic actions when we do something by first doing something else. But when this is not the case, there is no basic action. The issue turns, at least in part, on the deep question of what it takes for an event to count as an action. Causal accounts of this issue tend to require a special class of basic actions, while some other sorts of account (e.g., neo-ascriptivist) tend not to.

⁹⁹ Franklin, *A Minimal Libertarianism*, 112.

¹⁰⁰ Franklin, *A Minimal Libertarianism*, 108.

(i) requires agents to possess cognitive and volitional abilities that allow them to be normatively competent and have reflective self control to exercise free will. (ii) posits that directly free actions are basic actions, while (iii) posits the role of agents in exercising free will is exhausted by the causal role of their mental states, such as reasons and desires. According to (iv), free agents need the ability to refrain from exercising their abilities implying that they must be able to exercise them in more than one way.¹⁰¹ Franklin holds that this account is not susceptible to the same problems as Kane's account. By not postulating efforts of will and not conceding that indeterminism diminishes control, libertarianism is placed onto a more promising track. The problems of luck and enhanced control are much more solvable on this account.

Franklin's account of free will is concerned with an account of freedom that is interested in free basic actions. To summarize the view, it seems fair to say that Franklin's view takes indeterminism as a necessary condition and adds a compatibilist friendly criterion (normative competence) and then stipulates that the primary focus of the account of acting freely is the idea of a basic action. Franklin's theory claims that an agent makes a free and morally responsible choice if their action is undetermined and meets appropriate compatibilist criteria.

Franklin raises legitimate concerns with the Kanean approach but they are not, I think, decisive. Franklin takes issue with Kane postulating efforts of will and the role he gives them, but I think his idea of basic action is much more problematic than efforts of will. Franklin takes it for granted that basic actions are not contentious when they are a topic of debate in philosophy of action.¹⁰² We do not need to say that efforts of will are basic actions in order to say that agents make them when they make SFAs. An agent makes an effort of will when there is an end, goal or

¹⁰¹ Franklin, *A Minimal Libertarianism*, 108.

¹⁰² Andrew Sneddon "Does Philosophy of Action Rest on a Mistake?" *Metaphilosophy* 32, no.5 (2001); Annette Baier, "The Search for Basic Actions," *American Philosophical Quarterly* 8, no. 2 (1971).

purpose they are trying to bring about and the motivation to bring about a particular end is one internal to the structure of an agent's will.

Franklin's dilemma can be solved by tackling the first horn. It is true that responsibility efforts for present SFAs can be traced back to all the way to the first SFAs of childhood which are not like all the other SFAs in our lifetime. The reason they are different is that there is no backlog of self-formed character which the agent is responsible for making. The agent has only agency in a compatibilist sense. Contrary to appearances I do not regard this as a problem for Kane. It is actually a virtue of his theory that libertarian freedom is an ability we are born with but something that needs to be used to develop it and it is only utilized after a certain moment of psychological development. Specifically, for persons to develop libertarian free will their psychological lives have to be complex in the sense that their wants and desires receive feedback in the form of resistance to their motivations. Consider a speculative case of an eight year old child who wants to take more than their fair share of cookies but their parent prohibits them. The child has compatibilist agency as they are (albeit in a limited way) competent enough to respond to their parent's imperative. Suppose this is the child's first SFA. They want to listen to their parent's command, but they also want to take more cookies and they debate with themselves if they should listen or if they should act contrary to their parent's wishes. Whatever they decide, they are beginning to form their character; their situation is one where they are torn between listening to authority and disregarding it in favor of oneself and one's wants and desires. They are beginning to develop libertarian agency from a compatibilist one.

Kane's businesswoman faces a conflict between competing demands, which are internal to her will, that motivate her to act, and whatever she decides she will be the originator of the goal, purpose, or end she chooses. The nature of the businesswoman's SFA involves who she becomes

through what she decides to do. SFAs involve conflict between immediate goals but also involve more besides. The cases involve moral and prudential conflicts between motivations internal to one's will. So, while just the choice of gelato is uninteresting for free will, an agent who has undertaken a strict diet and is tempted to cheat on their diet with gelato is an appropriate scope. For such a conflict could be a prudential SFA where agents decide to shape their wills in such a way as to be the kind of person who chooses immediate gratification over a commitment to a long-term goal like their health.

e) *Freedom as Undefeated Preference Authorization*

Ekstrom's account has developed and changed from her earlier (2000) account to her more recent (2016) statement of her position. On her refined view, free will is defined in the following way:

Directly Free Action as Action Indeterministically and Non-deviantly Caused by Reasons of the Agent's Own: A decision or other act is directly free just in case it is caused non-deviantly and indeterministically by reasons of the agent's—such as convictions, desires, values, belief, and preferences—and other reasonable compatibilist conditions on free action are met, including that the act is not compelled and is not the result of (non-self-arranged) manipulation or coercion. An agent's performing a directly free act requires that it be open to her at the time not to perform that action, either by performing an alternative act right then or by not performing any action at all right then.¹⁰³

Let's attend to two technical notions from Ekstrom: "preferences" and "acceptances."

She says, "The term "preferences" I use not in the ordinary sense, but instead as a term of art, to refer to desires one forms or maintains as one aims to desire what is good."¹⁰⁴ She states elsewhere, "Call a desire—to act or to intend or to desire in some particular way—that has

¹⁰³ Ekstrom, "Toward a plausible event-causal indeterminist account of free will," *Synthese* 196, no. 1 (2019): 137.

¹⁰⁴ Ekstrom, "Towards a plausible event-casual indeterminist account of free will," 132.

survived ... a process of critical reflection with respect to the good a *preference*.”¹⁰⁵

“Acceptances,” are beliefs that the agent reflected on and came to see these beliefs as true.¹⁰⁶

A third important piece of her account is the concept of “undefeated authorization”,

df, undefeated authorization: *S* is authorized in preferring *A* in a way that is undefeated if and only if *S*’s evaluative faculty was neither coerced nor causally determined by anything to form the preference for *A*, but rather the preference for *A* was indeterministically caused by *S*’s considerations.¹⁰⁷

The idea of an undefeated authorized preference is understood as the consideration that ends up being indeterministically caused by what occurred to *S* at the time. Because it’s

indeterministically caused, *S* could have formed a preference other than she did holding fixed the same past and the laws of nature. An example Ekstrom uses to illustrate her view is taken from Charlotte Bronte’s classic novel *Jane Eyre*.¹⁰⁸ Jane Eyre discovers her beloved Mr. Rochester is still married and becomes divided over whether to stay or leave. Ekstrom describes the case as so:

The possible outcomes to her deliberation are the preference to remain with Mr. Rochester and live as his wife, the preference to remain yet to live as his platonic companion, and the preference to flee the situation altogether. Various considerations occur to her, including the thought that Mr. Rochester is intelligent and passionate, a desire for romantic love, a preference for lifelong companionship with someone who shares her interests, an acceptance that it is immoral to be romantically involved with a married man, a preference for acting in accordance with the divine will, an acceptance that it is the divine will for her to lead a morally upright life, a preference for having children of her own one day, a conviction that children are best off being born to married parents. Each of these deliberative inputs points rationally toward one or another of the

¹⁰⁵ Ekstrom, “Indeterminist Free Action,” 144.

¹⁰⁶ Ekstrom, “Indeterminist Free Action,” 144.

¹⁰⁷ Ekstrom, “Indeterminist Free Action,” 146.

¹⁰⁸ Charlotte Bronte Jane Eyre. In *The Selected Works of the Bronte Sisters*. (Hertfordshire: Wordsworth Editions, 2005).

differing conclusive preferences, and each raises the probability of one of the three potential deliberative outcomes.¹⁰⁹

Ekstrom takes the combination of an agent's acceptances and preferences to be constitutive of their character. She then takes the self to be a character combined with a capacity of critical reflection to form or reform that character. She writes, "the agent's self or identity is constituted by this faculty for critical reflection with an eye toward truth and goodness, together with the preferences and acceptances formed as the outcome of that reflection."¹¹⁰

Here is where I find an overlap between Kane and Ekstrom's accounts of freedom. Ekstrom believes that the proper location of indeterminism and alternative possibilities when freely acting is in how our "selves or characters" develop and we become the kinds of persons that we are. She says, "I came to think... that what was important is that we could have had different selves or characters than we did. What counted was that it was not rigged from the start who we turned out to be. Our actions should flow from our characters in order for those actions to be self-determined, and we should have a chance to be other than the sorts of people we are. The past and laws had to leave some openness in this matter."¹¹¹ Ekstrom's account is concerned with how we became the kind of person that we are. The notable difference is that the actions by which we become who we are not a unique kind of action like Kane's SFAs. Instead, the focus is on preferences and how those preferences are freely made and in turn cause agents to act.

Preferences are at the same time in Ekstrom's account the focus of her analysis and the target of objections. I will deal with the objections I think relevant to the discussion at hand. The

¹⁰⁹ Ekstrom, *Indeterminist Free Action*, 147.

¹¹⁰ Ekstrom, "Indeterminist Free Action," 144.

¹¹¹ Ekstrom, "Towards a plausible event-causal indeterminist account of free will," 133.

first objection comes from Randolph Clarke who argues that Ekstrom's account is vulnerable to not leaving free action *up to the agent*.¹¹² Ekstrom's (2000) account allows the formation of an intention to act be causally determined by one's reasons, particularly one's undetermined preferences. The account claims that holding fixed both the prior deliberation and the formation of preference, an agent could have done otherwise than form that decision to act.¹¹³ Ekstrom's older view believed that it was up to the agent what preference they would have formed because the preference would have been formed for reasons of the agent and indeterministically caused by those reasons without coercive influence. The agent has the categorical power to do otherwise than form the preference they did. The problem that this creates according to Ekstrom is, "there is no single act that is both self-determined and could have been otherwise."¹¹⁴ Ekstrom (2000) believed this to be a strength because she held that it constructs a direct flowing connection between the self and actions and that it allowed that we could have become different sorts of people because "person" is understood as the sorts of characters we have.

Ekstrom has changed her mind slightly with regard to allowing intentions to be passively indeterministically formed. She now holds that intentions, when made for and caused by reasons of the agent in deciding, "are not passive states but are actively formed. And intention formations that were indeterministically caused by the agent's formation of certain preferences would not be accidental or lucky."¹¹⁵

¹¹² Randolph Clarke, *Libertarian accounts of free will*. (Oxford: Oxford University Press, 2003).

¹¹³ Ekstrom, "Towards a plausible event-causal indeterminist account of free will," 135.

¹¹⁴ Ekstrom, "Towards a plausible event-causal indeterminist account of free will," 135.

¹¹⁵ Ekstrom, *Free Will*, 105.

Derk Pereboom argues Ekstrom's earlier account is vulnerable to the following problem,

...an infinite hierarchy of preferences is impossible for us, and so we must at some point arrive at fundamental preferences. On Ekstrom's event-causal picture, such fundamental preferences can't themselves be freely formed in the sense required for moral responsibility for them, since any such formations can't be backed by preferences. Thus the agent will not be morally responsible for the fundamental preferences, and also not for any preference formation or decision by them, and this undermines the position.¹¹⁶

Pereboom's regress objection conjoined with Clarke's objection convinced Ekstrom that she needed to fix her view. Her revised view, Directly Free Action as Action Indeterministically and Non-deviantly Caused by Reasons of the Agent's Own, is seen by Ekstrom to be able to answer the objections raised by Clarke and Pereboom.

Ekstrom holds firm that her later account of free action establishes our actions as "up to us" and her later account has a notion of up-to-us that utilizes alternative possibilities. She also thinks this has the ability to avoid any charge of luck, randomness or arbitrariness. This is the case because agents on this model self-determine their actions by having their reasons—like their values, beliefs, desires—indeterministically and non-deviantly cause themselves to act without coercive influence.¹¹⁷

Ekstrom's model has components in its favor that are shared by Kane. The particular element I have in mind here is her concern with how persons become who they are; persons have to be involved in the making of themselves through undetermined choices where it is metaphysically possible to choose otherwise. She, like Kane, does not appeal to extra factors like irreducible substances with special causal powers. There must be room *within* our environment for human choices, purposes, goals, decisions, and aspirations to make some kind of difference

¹¹⁶ Derk Pereboom, "The disappearing agent objection to event-causal libertarianism" *Philosophical Studies*, 169, no.1 (2014): 59-69.

¹¹⁷ Ekstrom, "Towards A Plausible Event-Causal Indeterminist Account of Free Will," 140.

in how we become the kind of person we become. In other words, nature itself must not have the final say on the kind of person we are but leave some space for human agency to play a role in this regard.

I am not convinced, however, that her model is preferable to Kane's account. Ekstrom's account takes the requirement of 'being up to the agent' quite seriously and she thinks that her refined view of directly free action secures this without being open to worries about luck. She thinks her view is preferable to Kane's model because of the location of the indeterminism. Ekstrom places the indeterminism before an agent acts so that her reasons indeterministically cause her action. By being able to cite an agent's reasons in causal explanations she believes we have an explanation of action that grounds our freedom and moral responsibility.

Ekstrom's account, I think, does not secure the direct link to free will and moral responsibility she thinks it does. Ekstrom's view of a directly free action places the indeterminism in between reasons and actions. But I think Kane's location of indeterminism in the production of action itself is the right place for libertarians to require it. On a libertarian view that has indeterminism in the reasons but not actions, then once an agent has indeterministically decided what to do the indeterminism exits the scene and the agent is no longer able to do otherwise at the time of action. This model of indeterminism has the decision process prior to the choice made and the way an agent decides is indeterministic while the action is determined. If we take away the indeterminism, then this view seems quite favourable to a compatibilist interpretation of freedom. The view seems to share an overlap with Harry Frankfurt's account of hierarchical free will (discussed in chapters 3 and 4) as freedom for our actions seem to lie in our ability to critically reflect on our motivations. The only difference

seems to be determinism as Frankfurt allows our reasons to deterministically cause our actions while Ekstrom requires our reasons to indeterministically cause our actions.

On Kane's account the indeterminism is present throughout the production of action as he views trying to decide what to do as an act itself. Since making a decision is an action it is something we do and since we are trying to do both actions, to settle our minds on one goal, we have responsibility whatever the outcome. The indeterminism is in the act of making a decision and so we could fail to do what we are trying to do but this does not undermine our control or responsibility. As I have argued just because an act is undetermined it does not follow that we have no control or responsibility for our actions. In the case of SFAs where we are trying to do multiple things, we are responsible no matter what the outcome even though it is undetermined. The virtue of this model of indeterminism is that, because we can fail at what we are trying to do, if we succeed at what we are trying to do we clearly are responsible because it is something we did. In other words, we can see that when an agent acts on this model of indeterminism it is quite clearly something they intentionally do and not accidentally. I think having the indeterminism at the moment of action is preferable because I think it is clearer that the agent really could have done otherwise because they were not resolved over what to do. Recall Kane's businesswoman. She could not use indeterminism as an excuse as to why she was late to her meeting or why she failed to help the assault victim. It would be a poor reason to say, "I didn't fail to help the victim, chance did." What this move is supposed to do is to not only make the businesswoman the ultimate source of her actions but also of her will. She is the ultimate source of her will because she directly shapes it through whatever choice she makes and is trying to make. Because she is trying to accomplish a particular goal and resist the temptation of the other goals, she has a direct role in shaping her will because she makes a motivation internal to her

will move her to act. This gives the Kanean model an advantage over Ekstrom's in my view because this model offers a model of agency where we shape our will in such a way to shape the reasons for why we act. Of course, this shaping should not be read as saying we have total control over our wills. Our wills are shaped by a complex web of causes which we do not have control over. The way we shape our wills through the choices we make like the ones in SFAs is in having our choices play a direct role in being what our wills are; had we not made a particular choice then our wills would not be what they are. Ekstrom's model does not secure the same level of responsibility for our wills that, I think, libertarian freedom should deliver.

I think we have a more direct level of responsibility on the Kanean approach because there is indeterminism right up until we choose. We make our minds settled then and there by choosing which happens at the same time as deciding. This makes action as undetermined as it could be making the responsibility for it the greatest it could be regarding responsibility for undetermined actions.

5. Sourcehood and Ultimacy

As I discussed in an earlier section Kane's view is an incompatibilist sourcehood view that says people are, "the ultimate *creators* (or *originators*) and *sustainers of their own ends or purposes*." (emphasis not mine)¹¹⁸ We are creators of our own ends or purposes in virtue of forming our wills through SFAs. SFAs are moments where we shape our wills in ways both seemingly trivial and more significant. We make SFAs about prudential matters as well as moral ones. The benefit of locating self-formation in choices with specific features is that only these choices are ones that require we be able to do otherwise. We still make free choices even when

¹¹⁸ Kane, *The Significance of Free Will*, 4.

they are not SFAs. This allows libertarians to concede that we do not need alternative possibilities for every single choice we ever make. I view this as a dialectical advantage over the alternative theories surveyed because we do not have to offer a view that I think would be too strong in its criteria for free will. As I understand them, the accounts surveyed would only ascribe free and responsible choice to undetermined actions while Kane would attribute it to determined and undetermined actions. This is seen in the schema describing free action and free will earlier in the chapter.

In my view, libertarians lack an account of authorship. Kane and Ekstrom both share an overlapping concern for how persons become the kind of people they are with the character, motives and wills that constitute who they are. I think that Kane's account is significant, particularly his account of self-formation but, I think, his account lacks a fully adequate account of the larger structure in which we make SFAs. I propose to develop and expand on Kane's view by demonstrating that free agents are the authors of their own life stories and themselves. The reason I think his account is insufficient without an account of authorship is that we lack a complete understanding of how and why we choose to become the kinds of persons we are in moments of self-formation. If we understand the businesswoman as an (metaphorical) author, we can better understand the choice she makes in the moment of self-formation. I also think we need an account of authorship because there is a clear concern over our personal identities in views like Kane's and Ekstrom's. Kane's and Ekstrom's accounts are clearly motivated by the concern that one's personal identity is predetermined and that persons are never free to choose to become who they are. Their worry seems to be that if determinism is true our identities are fixed because

they are the products of necessitating influences we cannot control.¹¹⁹ This is the topic of chapter four. I discuss in chapter four the insufficiency of compatibilism stemming from the absence of a satisfactory view of authorship. In chapter five I will delve into my account of authorship and the relationship between authorship and agency. In the next chapter, I am going to critically discuss compatibilism. I will provide an overview of classical and more recent compatibilist theories, and then go into my critical discussion of them. This will lay the groundwork for chapter four and the argument in that chapter for incompatibilism.

¹¹⁹ Kane says something about the first SFA in childhood that is relevant here. He says the very first SFA can be determined. A critic might say it is fine for every SFA to be determined. I show in chapter four that this is not true. In chapter four, I discuss how libertarianism is attractive compared to other options because libertarianism delivers an open future, and every SFA being determined would threaten this.

Chapter 3

Compatibilism

1. The Compatibilist Orthodoxy

The last chapter provided readers with an outline of libertarian theories and analyzed Kane's view and the challenges to it. The main takeaway from that chapter was that among the libertarian options, a broadly Kanean framework is the right approach to the issue of free will and moral responsibility. The chapter was largely exploratory in surveying various libertarian theories although it was argued that Kane's view is the best game in town. I now turn my attention to the task of engaging with a direct challenge to libertarianism: compatibilism. Recall from the last chapter the metaphor of "incompatibilist mountain." We begin our ascent up incompatibilist mountain and our journey up the mountain will be done over the next two chapters. This chapter will provide a critical discussion of compatibilism while in the next chapter I argue for incompatibilism. This chapter sets the stage for the argument in chapter four by arguing in a negative sense that compatibilism is not a correct account of free will and moral responsibility. We need to take care when dealing with the present topic as there are as many compatibilist theories as there are libertarian ones. Moreover, since I don't dismiss compatibilism wholesale in the type of analyses found within the label, what is wrong with the doctrine needs to be precisely articulated. I also want to show the aspects of various views that I think are involved in a correct view of free will and moral responsibility. The discussion to be carried out will argue that these aspects are all pieces of the larger free will puzzle but they are not the whole picture.

I will begin with identifying the classical and contemporary strategies philosophers have articulated in outlining compatibilism. Then I will begin my argument for why these views fail to

deliver to agents the ability to be an author of their own life story. The way I will go about doing so is to discuss the relevance of history for free will and moral responsibility. The reason for this is it will allow us to separate the kinds of compatibilism on grounds that show in exactly what way I believe compatibilism is deficient. The first section will be divided into an overview of compatibilist theories while the second section will introduce the manipulation argument. I will give the argument in its general form and then situate the argument in terms of each of the theories presented in the first section. Through bringing in this argument we can evaluate compatibilism on similar grounds to my own incompatibilist theory. This provides the set up for the discussion in the third section where I argue in exactly what sense I find compatibilism lacking. This will provide a springboard into the next chapter where I present my view of libertarian free will.

2. Compatibilism: Classic and Contemporary

a) Classical Compatibilism

For introductory purposes I will start with the compatibilist view I identified in the introductory chapter: generic compatibilism. This is the view that free will and moral responsibility are compatible with causal determinism. The history of philosophy is rife with examples of philosophers who have a passionate dislike of generic compatibilism. For instance, William James called it a “quagmire of an evasion,” and Immanuel Kant described it as a “wretched subterfuge.”¹²⁰ These disparaging comments aside, generic compatibilism enjoys an

¹²⁰ William James "The Will to Believe." In *The Will to Believe and Other Essays in Popular Philosophy* (New York: Longmans, Green, and Co., 1897): 149; Immanuel Kant, *Critique of Pure Reason* trans. Paul Guyer, and Allen W Wood. (Cambridge: Cambridge University Press, 1998): 78

orthodoxy in contemporary philosophy.¹²¹ This section aims to unpack generic compatibilism and illuminate the differing views that philosophers from ancient times to the more contemporary have articulated. I begin with what we can call “classical compatibilism,” so named because the form of the argument it takes, as I will show momentarily, can be traced back to some ancient thinkers. Classical compatibilism asserts that the nature of freedom is to be understood in terms of an opposition to constraint and constraint is understood in terms of the constraint of other agents who use coercion to interfere in your life. Freedom is described as having two conditions: firstly, it is the power or ability to obtain or fulfill whatever you want or desire you have, and secondly, no other agent constrains you through coercion from obtaining or fulfilling your wants or desires. An example of a constraint that undermines freedom is an agent that threatens another agent with a gun to hand over their wallet. Acting to comply with the threat is not free, by the standard of the position being considered. Freedom is thus understood negatively as it is the absence of constraints like coercion as well as any compulsion.¹²² Ancient thinkers like the Stoics and modern philosophers like John Locke, Thomas Hobbes, and David Hume all believe freedom just is the absence of constraints like the one described and that this conception of freedom is compatible with determinism.¹²³ For these thinkers determinism is not a constraint on our freedom because it does not stop us from fulfilling our desires. As we saw in

¹²¹ At the time of writing the PhilPapers survey reports 58 percent of professional academic philosophers surveyed either lean towards or accept compatibilism. Now this data set is generalist and perhaps among the relevant subfields like the philosophy of action or mind it might differ but without such data at hand I take it to be the accepted view.

¹²² Kane, *A Contemporary Introduction to Free Will*, 13.

¹²³ Kenny, *A New History of Western Philosophy*, 82; John Locke, *An Essay Concerning Human Understanding* (Oxford: Oxford University Press 2008): bk.II, ch. XXi; Thomas Hobbes, *Leviathan*. eds Crawford Brough. (London: Penguin, 1985 1958): Pt. II, chap. 21; David Hume *A Treatise of Human Nature* eds Ernest Campbell (Mossner. Harmondsworth: Penguin, 1985): section VIII; John Stuart Mill *On Liberty*. Eds. Currin V. Shields (Indianapolis: Bobbs-Merrill, 1956); James P Scanlan, “John Stuart Mill and the Definition of Freedom.” *Ethics* 68 (1958): 194-206.

chapter two, determinism holds that there is only one possible future and so we cannot do otherwise than we actually do. But this is of no concern to classical compatibilism because even if we are determined to act as we do what matters is that we are doing what we want to do.

Consider, as an example, the comparison that the “second father” of Stoicism Chrysippus makes between human agents and a dog tied to a cart: “When a dog is tied to a cart, if it wants to follow it is pulled and follows, making its spontaneous act coincide with necessity ... So it is with men too...”¹²⁴ Chrysippus appears to be saying that an action can be made spontaneously and necessarily at the same time. An action that is done out of spontaneity is one that is done freely. So, though the dog moves necessarily because it is being dragged, because it moved spontaneously at the same time it still moved freely. The way Chrysippus is comparing human agency to the dog in the passage is that an agent can act both necessarily and spontaneously i.e., be determined to act and act freely. The way that an agent acts freely while being determined to act is that, like the dog in the passage, the agent acts in accordance with their wants. While not in the passage itself, it seems fitting to say that the negative view of freedom described in the last paragraph applies. As long as we do what we want to do, that we also act of necessity is beside the point.

Hobbes describes this type of freedom like so:

A Free-man, is he, that in those things, which by his strength and wit he is able to do, is not hindred to doe what he has a will to... Liberty and Necessity are consistent: As in the water, that hath not only liberty, but a necessity of descending by the Channel: so likewise in the Actions which men voluntarily doe; which (because they proceed from their will) proceed from liberty; and yet because every act of man’s will, and every desire, and inclination

¹²⁴ A. A Long and D. N. Sedley, *The Hellenistic Philosophers*. Vol. 1: *Translations of the Principal Sources with Philosophical Commentary* (Cambridge: Cambridge University Press 1987): 62a.

proceedeth from some cause, and that from another cause, which causes in a continuall chaine... proceed from *necessity*.¹²⁵ (author italics)

Hume offers a very similar definition of freedom in his *Treatise of Human Nature* where he makes a distinction between two kinds of freedom common at the time: the liberty of spontaneity and the liberty of indifference. He associates acting freely with the former, writing:

Few are capable of distinguishing betwixt the liberty of spontaneity, as it is call'd in the schools, and the liberty of indifference: betwixt that which is oppos'd to violence, and that which means a negation of necessity and causes. The first is even the most common sense of the world; and as 'tis only that species of liberty, which it concerns us to preserve, our thoughts have been principally turned towards it, and have almost universally confounded it with the other.¹²⁶

Locke's analysis of freedom is also articulated along these lines. Freedom, which he interchanges with liberty, is "the *Idea* of a Power in any Agent to do or forbear any particular Action, according to the determination or thought of the mind, whereby either of them is preferr'd to the other..."¹²⁷(emphasis original). Locke argues that the problem of free will and determinism should be understood not as a question of the freedom of the will but rather in terms of the freedom of the agent. He says this explicitly writing that the very question of the freedom of the will is "as insignificant to ask... as whether his Sleep be Swift, or his Vertue Square, *Liberty*, which is but a power, belongs only to Agents, and cannot be an attribute or modification of the *Will*"¹²⁸ (emphasis original). On Locke's view, the problem of determinism and free will is properly seen as a problem of determinism and free agency, and as he sees it they are compatible. For free agency requires that, "so far as any one can, by the direction or choice of his

¹²⁵ Hobbes, *Leviathan*, 261-262.

¹²⁶ Hume, *A Treatise On Human Nature*, 455.

¹²⁷ Locke, *An Essay Concerning Human Understanding*, 143.

¹²⁸ Locke, *An Essay Concerning Human Understanding*, 145.

Mind, preferring the existence of any Action, to the non-existence of that Action, and, *vice versa*, make it to exist, or not exist, so far he is *free*.”¹²⁹ Locke offers the following example to illustrate his point:

Suppose a Man be carried, whilst fast asleep, into a Room, where is a Person he longs to see and speak with; and be there locked fast in, beyond his power to get out; he awakes, and is glad to find himself in so desirable Company, which stays willingly in, *i.e.* prefers his stay to going away. I ask, Is not this stay voluntary? I think, no Body will doubt it; and yet being locked fast in, ‘tis evident he is not at liberty not to stay, he has not freedom to be gone.¹³⁰

The examples given of Chrysippus, Hobbes, Hume, and Locke show a common thread in their philosophical analysis of freedom. To be free just is to act in accordance with your wants and desires. If your wants and desires be determined and render you unable to do otherwise, this is of no concern to this conceptualization of freedom.

To be clear, this line of thinking does have something to say about the freedom to do otherwise and we can derive it from the passages we have already seen. The freedom to do otherwise is a power or ability because freedom is a power or ability, and it is a conditional or hypothetical one. What I mean is that if freedom is the ability to do what one wants, then one can do otherwise if one had wanted to. Consider the following case of deciding to eat a piece of cake or an apple. A conditional analysis of freedom like that found in classical compatibilism would say something like

Freedom of the will (or free agency) means that you are free to choose to eat the cake (1) if you have the power or ability to eat the cake and, (2) no one is forcing you to eat the

¹²⁹ Locke, *An Essay Concerning Human Understanding*, 148.

¹³⁰ Locke, *An Essay Concerning Human Understanding*, 144.

cake or constraining you from eating the apple. If you eat the cake, you could have eaten the apple if you had wanted to instead of wanting the cake.

The conditional operating in this paragraph is of the form “would...if”. You would have acted otherwise if you had wanted to. Other things being equal, I choose the cake but if things were not equal then I would have decided to take the apple.

An updated form of this kind of compatibilism is offered by Daniel Dennett who goes even farther and says we have no need of alternative possibilities in any sense.¹³¹ Dennett uses two cases to make this point: the first is Martin Luther and his utterance “Here I stand I can do no other.” The second case he offers is one involving himself being offered a million dollars to torture someone and being unable to do anything other than refuse.¹³² Dennett believes that the question of “could he have done otherwise” is mostly forward-looking. There is a backward-looking component to it as actions can sometimes express interesting features about the agent such as their character.¹³³ We use this backward consideration in a forward-looking way by reflecting on how we might change our behaviour if that is warranted. So, if I reflect on how my constant overconsumption of dessert reveals I am gluttonous, I can take that information to try to reform myself. In a positive sense, if I see that I am really good at a particular activity then looking into the future I might consider sustaining that pursuit.

Dennett is not successful, however, in establishing there is no deeper significance to the backward-looking aspect of “could have done otherwise.” Dennet describes the ability to have

¹³¹ Daniel Dennett, *Elbow Room: The Varieties of Free Will Worth Wanting* (Cambridge, Massachusetts: MIT Press, 2015): 145.

¹³² Dennett, *Elbow Room*, 145.

¹³³ Dennett, *Elbow Room*, 148.

done others as being able to modify one's behavior, but there is more to the story than just modification. A fellow compatibilist Gerald Dworkin believes that Dennett's attempt to make could have done otherwise purely consequentialist fails. He says that, "I believe that any attempt to forge as close a link between responsibility and modifiability as Dennett wants ignores those ascriptions of responsibility which are not oriented toward the future, but are, so to speak, for the record. And since they are for the record, justice requires that we pay attention to the details of a person's circumstances, and not to what is true in general or true of individuals very similar to her."¹³⁴ Dennett focuses on the beneficial consequences of "accepting" responsibility for one's actions. His view of responsibility is quite consequentialist and pragmatic, but it ignores kinds of responsibility that are about moral desert. I take Dworkin to be making this very point as there are cases of responsibility that are about deserving praise or blame for their actions. Dworkin uses the case of sleepwalking to show a sense of responsibility that is separate from a consequentialist one: "It might have good consequences to hold sleepwalkers responsible for what they do in the sleepwalking stage because otherwise too many people would pretend to have been sleepwalking. But one does not think they really are responsible. So the linkage between being responsible, being capable of being modified by sanctions, and having good consequences result from being held responsible is not as tight as Dennett would have it."¹³⁵

Just attributions of responsibility take into account where appropriate a person's circumstances, which will include their history. A proper assessment of an individual's circumstances will be how they got to be the way they are. This is relevant to assessing a person's competency as personal history will provide needed information to properly judge if

¹³⁴ Gerald Dworkin, "Elbow Room Daniel C. Dennett," *Ethics* 96, no.2 (1986): 424.

¹³⁵ Dworkin, "Elbow Room Daniel C. Dennett," 424.

one is competent. Consider the following case of a young man at trial. The case is a real trial that Kane attended and witnessed.¹³⁶ The young man was on trial for the rape and murder of a young girl in the local community. Kane describes his initial reaction of resentment and anger towards the young man because he and his family were friends with the victim's family. But as Kane listened daily to the testimony of how the young man came to have the mean character and perverse motives he did have—the result of parental rejection, sexual abuse, bad role models and other such factors—some of Kane's resentment toward the young man decreased and shifted to the young man's abusers. However, Kane says that he was not ready to shift all the blame away from the young man.¹³⁷ The formative circumstances of the young man's circumstances are important to take into account as they play a role in the young man's responsibility. Explanations of actions that are intended to provide the grounds of judging blameworthiness or praiseworthiness in a fair way will involve going beyond someone making a choice to include why they made that choice. After all, if one chooses to make oneself in such a way that is abnormal and results in direct harm to others, this cries out for a further explanation. Why did the young man decide to become such a cold-hearted person inflicting cruelty to others? The answer to this will involve personal historical circumstances to provide an explanation that fairly weighs his moral desert for his behavior. Simply citing a decision was made is not enough for a fair judgement of his accountability. If we can trace his action to a choice and then trace that choice beyond just his decision, this is a relevant point to consider at his trial. It might not fully excuse him, but it is a point that must be weighed for a complete accounting of his moral blameworthiness.

¹³⁶ Robert Kane "The Problem of Free Will: A Libertarian Perspective," In *Do We Have Free Will?: A Debate*, eds Robert Kane and Carolina Sartorio (New York: Routledge, 2022): 27

¹³⁷ Kane "The Problem of Free Will, 27.

Now this does not mean the young man could not be punished for consequentialist reasons like the possibility he might kill again. But as regarding the status of the young man as deserving of blame, it seems that he does not deserve to be punished when it was causally impossible for him to ever make himself different than he was. This seems to me to align with the intuition that there are circumstances of birth and influences of upbringing, things like mental disabilities and child abuse, that can limit free will and moral responsibility. This intuition is good reason to be concerned not only with whether the young man is sane or insane but the circumstances around why he chooses to become who he is.

b) Deep Self Compatibilism

I am going to discuss the contemporary forms of compatibilism by using common themes to group them together so as to cast as wide a net as possible while still keeping the scope manageable. The first class of views we will consider is referred to as “Deep Self Compatibilism” by Susan Wolf.¹³⁸ These views share the characteristic that the kind of moral responsibility thought to be related to free will is found in the responsibility for the self that performs those actions.¹³⁹ In this way, these accounts recognize that the account of freedom in Hobbes and his contemporaries and others is not a complete account. As we will see momentarily, Wolf, Frankfurt and Watson spell out a compatibilist account of freedom of will, not just freedom of action.

i) Frankfurtian Deep Selves

¹³⁸ Susan Wolf, “Sanity and the Metaphysics of Responsibility.” In *Free Will*, eds Robert Kane (Malden, Massachusetts: Blackwell Publishers 2002): 148

¹³⁹ Wolf, “Sanity and the Metaphysics of Responsibility,” 148.

Frankfurt's account of free will is focused on endorsing the structure of one's will. To endorse the structure of one's will is to exercise an ability where we reflect on whether we want to act for one desire or another. Being responsible for our selves on Frankfurt's view requires only that one endorses the structure of one's will. Frankfurt's view of free will takes this to be the same as endorsing one's self as his paper is both an account of free will and personal identity. So, when Frankfurt talks about endorsing one's will, he can also be read as addressing one's self or the self one wants to be. There is an important distinction in Frankfurt's view between first-order desires and second order desires.¹⁴⁰ The desires of the first order have as their object of desire literal objects, actions, events and more: I want a chocolate milkshake. The second order desires have desires as their objects of desire: I want to want the milkshake.

The key piece of this view is Frankfurt's technical term of second-order volition. A volition is a higher-order desire for a certain desire to be effective in action and for it to be one's will.¹⁴¹ Imagine two drug addicts with symmetrical first and second order desires. Each of them possesses competing first-order desires to ingest their drug of habit and a desire to resist the temptation. Now, where the symmetry breaks down and we are able to pick out the one drug addict as a person from the addict that is a wanton is that the person has a volition about their desires. Persons want certain desires to move them to act while wishing other desires did not. The drug addict who is a person has a higher-order volition for his desire to not take the drugs to be the one that moves him to act. The wanton addict, however, cares not what desire moves him to act. Here is how Frankfurt carves out the difference between the addicts:

¹⁴⁰ Frankfurt, *The Importance of What We Care About*.

¹⁴¹ Frankfurt, *The Importance of What We Care About*.

The second of these addicts may suffer a first-order conflict similar to the first-order conflict suffered by the first. Whether he is a person or not, the wanton may... both want to take the drug and want to refrain from taking it. Unlike the unwilling addict, however, he does not prefer that one of his conflicting desires should be paramount over the other; he does not prefer that one first-order desire rather than the other should constitute his will. It would be misleading to say that he is neutral as to the conflict between his desires, since this would suggest that he regards them as equally acceptable. Since he has no identity apart from his first-order desire, it is true neither that he prefers one to the other nor that he prefers not to take sides.¹⁴²

The wanton addict does not care if he succumbs to temptation or not while the other addict wishes very much for his desire to resist to be the one that springs him to act. Through the technical notion of volition, Frankfurt's account develops a freedom of the will that is distinct from freedom of action. Wolf illustrates the distinction in Frankfurt's own addict case¹⁴³; one has the freedom of action if one has the freedom to do what one wills to do—the addict has the freedom to act on his desire to consume drugs. Freedom of the will is the freedom to will whatever it is one wants to will. What that means is that I am able to choose the desire that I want to choose. So, the unwilling addict lacks freedom of the will because they are unable to will what they want to will which is to not take the drugs. The unwilling addict's lack of freedom of will translates into not being responsible for his self as he is not responsible for his will. The fundamental reason the addict is not responsible is the desire he wants to move him to act is not the one that ends up motivating him to act.

ii) *Watsonian Deep Selves*

Watson like Frankfurt is sensitive to distinguishing freedom of action from freedom of the will. Watson's approach departs from Frankfurt's in its focus on values and desires. Both of their

¹⁴² Frankfurt, *The Importance of What We Care About*, 18.

¹⁴³ Wolf, "Sanity and the Metaphysics of Responsibility," 150.

accounts have an overlapping concern (albeit in different ways) with the structure of an agent's will and not just with the conditions of action for evaluating the existence of free will. Watson's account can be cast in terms of selves by paying attention to the connection between selves, desires and values. When our desires align with what we value our action expresses our whole self.

Watson engages the problem in the following way: "The problem of free action arises because what one desires may not be what one values, and what one most values may not be what one is finally moved to get."¹⁴⁴ I may value a commitment to long term health but I have a short term desire which conflicts with this and which could be the moving principle of action. When our values and desires come apart in this way and we act contrary to what we value, we do not will freely. When our behavior does not have this alignment, our selves do not align with our values. Values for Watson are:

those principles and ends which (one)—in a moment of cool and non-self-deceptive reflection—articulates as definitive of the good, fulfilling, and defensible life. That most people have articulate "conceptions of the good," coherent life-plans, *systems* of ends, and so on, is something of a fiction. Yet we all have more or less long-term aims and normative principles that we are willing to defend. It is such things as these that are to be identified with our values.¹⁴⁵

In judging whether an agent wills freely, we need to know if an agent's valuational and motivational systems overlap in the appropriate way. Agent's valuational systems are sets of considerations that in conjunction with one's beliefs take the form "I should A" The form this takes in action is "I think the course of action I should be taking is A." An agent's motivational system is a set of considerations understood as desires or motivations. "I want A" or "I want to

¹⁴⁴ Gary Watson, "Free Agency" *The Journal Of Philosophy*, 24, no.8, (1975): 205- 220.

¹⁴⁵ Watson, "Free Agency," 215.

do A.” The two perspectives of these systems can come apart as “I want to do A” will not always align with “I think I should do A”. Everyday life is filled with examples of cases where this happens. “I think I should exercise” and “I want to not exercise and watch television.” The former is an example of what I value and the latter is an example of what I want. Watson’s view of the problem of free will is that we can act contrary to what we value. He views it as “the fact that an agent’s valuational system and motivational system may not completely coincide. Only when an agent is moved to act by what they do value they act freely of their own free will.”¹⁴⁶ I would put it like so: to the extent that what one values expresses the sort of the person one, the alignment of motivations which sometimes also expresses the sort of person one is with our values, we act in accordance with our selves and act freely. Harry values his health and is motivated to adequately exercise and consume a well-balanced diet. The alignment of his values and motivations deliver freedom by the standards of the current position.

Watson’s view does not say we need a higher-order volition which we desire to motivate us to act and so he avoids a problem Frankfurt has to contend with. That is the issue of how we stop reflecting on our desires in those moments that we do pause and think about what we want. It seems theoretically possible that we could have an infinite series of higher-order desires and no way to cut off the reflecting. Watson’s view does not have this issue, but there is another problem which is related to the philosophical problem of *akrasia*. Briefly, *akrasia* is the phenomenon of acting contrary to what one judges to be the best course of action. If I judge exercise would be the best thing for me to do but I stay home in bed I have acted akratically. *Akrasia* is a controversial topic as some philosophers (the most notable being Socrates) deny its existence. I am going to assume it does exist for my purposes. On a Watsonian framework, it

¹⁴⁶ Watson, “Free Agency,” 215.

appears that agents do act akratically because akratic action is acting contrary to what you think you should do. I think I should go to the gym and exercise but I don't and instead stay home and eat pizza. Watson's view says this is an unfree act, but I think we have reason to be skeptical. Whether or not akrasia is a real phenomenon, I think there is a separate question of whether it should make our actions unfree. Do we freely act contrary to what we think we should do? I think it is at least possible. If we feel motivated to act by something we feel that is part of us, I don't see why this would exclude our action from acting freely. For example, keeping a promise is something I think I should do, but I could be motivated not to by other motives that are really important to me. I may have promised my wife to do some house chores and I think honouring this promise is what I should do, but I also really want to go out with some friends and relax.

Implicit in the present discussion is a distinction in the obstacles to one's freedom cast in terms of inner and external barriers in one's will. External barriers are rather obvious cases like being held at the end of a knife. Internal barriers are a bit more subtle as it might strike one as odd that one's own will could get in the way of one's freedom. With the help of a little reflection, we can see this actually quite plausible. The phenomenon of addiction seems to me to be the prime example of what Watson has in mind. An addict who is trying to achieve sobriety presumably values that goal, but their addiction obviously has cultivated a desire for a substance. On a classical view of freedom, ingesting the substance is free because one has a desire for it, yet it is precisely this desire that gets in the way of what one values, sobriety. It is not that the desire or substance is an obstacle because it is an alternative. It is an obstacle in exactly the sense of getting in the way of pursuing what one values: sobriety. Watson thinks that what makes akratic action undermine free will is that agents are "unable to get what he most wants, or *values*, and this inability is due to his own motivational system. In this case the obstruction to the action he

most wants to do is his own will. It is in this respect that the action is unfree: the agent is obstructed in and by the very performance of the action.”¹⁴⁷ It seems to me that if the nature of the unfreedom is one’s will then it seems fair to ask to what extent did one shape one’s will such that one cannot act in accordance with one’s values. Is the structure of one’s will the victim of bad luck or did an agent shape it through choice in prior cases where they faced a similar conflict? It seems that one could freely act akratically if the akratic structure of one’s will is at least partly a result of one’s prior choices.

iii) *Sane Deep Self Compatibilism*

Susan Wolf thinks there is a lot of insight into the “deep self” view in Watson and Frankfurt but she thinks they are incomplete. Wolf departs from Watson and Frankfurt about whether a will that is free needs only to be structured in accordance with the deep self’s deepest commitments. Her view brings in personal history in a way that Watson and Frankfurt do not. Wolf agrees with Watson and Frankfurt that, “responsible agency involves something more than intentional agency. All agree that if we are responsible agents, it is not just because our actions are within the control of our wills, but because, in addition, our wills are not just psychological states *in* us, but expressions of characters that come *from* us, or that at any rate are acknowledged and affirmed *by* us.” (author’s italics).¹⁴⁸ These remarks focus on the conditions of agency and more specifically the conditions for responsible agency. What is meant by responsible agency? Well, the most obvious to me is ‘morally’ responsible agency as we have been discussing the deep self and the requirement to be morally responsible for our deep selves according to

¹⁴⁷ Watson, “Free Agency”.

¹⁴⁸ Wolf, “Sanity and the Metaphysics of Responsibility,” 147.

Frankfurt and Watson. Wolf, Frankfurt, and Watson all consider free will to be relevant to moral responsibility. Wolf describes the relationship between free will and moral responsibility under the umbrella of the deep self through a particular ability: sanity.

Consider Wolf's case of JoJo, the son of a fictional ruthless dictator. JoJo is raised to follow in his father's footsteps, and he looks up to his dad as a role model and someone to emulate. He therefore embraces his father's values and when he grows up and takes over, he engages in the same ruthless behavior; he sends his fellow citizen to prison, to death or to be tortured on nothing more than the slightest whim.¹⁴⁹ Now Wolf sees a problem with attributing free will to JoJo because JoJo's personal history causally necessitated that he become a ruthless kind of person.¹⁵⁰ Furthermore, JoJo satisfies the conditions for freedom on the classical and Deep Self accounts, but that he does suggests something deficient with these views. Wolf proposes what is missing is the criterion of *sanity*. Sanity is an ability to know the difference between right or wrong either immediately or with reflection.¹⁵¹ What is wrong with JoJo on this picture is that he is insane. What constitutes his insanity is not a failure to use this ability but the absence of it. No matter what JoJo does, he will never be able to see that someone who fails to salute him does not deserve to be tortured.¹⁵²

According to Wolf, agents who are insane are unable to do the right thing for the right reasons and therefore cannot recognize the True and the Good.¹⁵³ Wolf goes on to make an

¹⁴⁹ Wolf, "Sanity and the Metaphysics of Responsibility," 153.

¹⁵⁰ Wolf, "Sanity and the Metaphysics of Responsibility," 151.

¹⁵¹ Wolf, "Sanity and the Metaphysics of Responsibility," 151.

¹⁵² Wolf, "Sanity and the Metaphysics of Responsibility," 154.

¹⁵³ Wolf, "Sanity and the Metaphysics of Responsibility," 158.

interesting claim about the compatibility of praiseworthiness and blameworthiness and determinism. It is her contention that our blameworthiness is incompatible with determinism. Her claim more specifically is that persons that are insane and so cannot do otherwise are not deserving of blame. They are determined to behave as they do and they are not blameworthy, so determinism is incompatible with blameworthiness. Praiseworthiness, however, is compatible with determinism. If one is sane and does the right thing, then even if one is determined to be sane and do good, you still deserve praise for the doing good. One case where this seems true is saving a drowning child and being psychologically determined such that it was not open to an agent they could not have done otherwise. This “asymmetry thesis” is a core component of the Sane Deep Self View. The basic idea is that determinism is incompatible with the total absence of the ability to know right from wrong and being determined to not have this ability makes it unfair to blame people. We can, however, be praised because we do have an ability (sanity) and the use of this ability is compatible with being determined to have it.

JoJo is not blameworthy because he was causally determined to be insane. Praiseworthiness is treated as purely structural and ahistorical in the sense that we only need to have an ability and it is not relevant how we come to have that ability. There are two considerations here which we need to keep separate. One is an empirical question: would we withhold blame from JoJo if we thought his behavior was causally determined but, at the same time, offer JoJo praise if he were a magnanimous ruler because he was causally determined? This is an issue for x-phi to probe. The other is a conceptual issue: Are the conditions for being blameworthy asymmetrical to the conditions needed to be praiseworthy? Wolf might insist that they are but since she invokes history to explain the exculpatory feature of insanity it can be used to probe her asymmetry thesis. Wolf’s view says that JoJo is insane because his upbringing involved factors beyond his

control that determined him to lack the requisite ability of sanity. But suppose that JoJo was not merely passive with respect to becoming who he was but played an active role in becoming the person he is. The way in which JoJo is determined is a relevant point. This seems like a relevant consideration since Wolf is the one who raises history and contends there to be a connection to moral blame.

Wolf's asymmetry thesis is partly correct as people's histories can influence them to such a degree as to mitigate their personal responsibility for becoming who they are and what they do. A case cited by Watson that demonstrates this is Robert Harris.¹⁵⁴ He was a brutal murderer that relished in killing innocent victims. Harris had a horrible upbringing filled with parental neglect and sexual abuse. For the sake of argument, suppose these experiences were so formidable that they determined Harris to become insane. We might still restrict his liberty for the safety of the community, but he is not to be seen as deserving to be blamed for his actions by Wolf's standards. But suppose for the moment that sometime in his life he purposefully chose to become a ruthless killer, to "harden his heart", resulting in losing his sanity. His decision determines him to lose the ability Wolf asserts is required to be blameworthy, but if he deliberately decides to determine himself in this way it seems he is, contrary to Wolf, blameworthy.

On the flipside of the coin, praiseworthiness is sometimes offered to people independent of history. A firefighter who saves a drowning child is rightly praised for their heroic deed without any inquiry into their history.

One final point I wish to make about Wolf's view is that it seems correct that the capacity to know the difference between right and wrong is an important one for moral responsibility and

¹⁵⁴ Gary Watson, "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme." In *Free Will*, edited by Robert Kane, (Malden, Massachusetts: Blackwell Publishers, 2002): 170.

free will. If we lack this capacity entirely, we cannot make a choice between right and wrong since we are incapable of knowing the difference. Where I have departed from Wolf is in determinism being irrelevant to praise and excusing blame irrespective of the choices one makes which shape oneself in important ways.

c) *Semicompatibilism*

Another view comes from John Martin Fischer: *semicompatibilism*. Semicompatibilism meets incompatibilism half-way by conceding that free will, understood as an ability to do otherwise, is incompatible with determinism. Fischer believes that Peter Van Inwagen's Consequence Argument is sound, and this argument rules out agents ever being able to do otherwise.¹⁵⁵ Recall that Van Inwagen's Consequence Argument says that if determinism is true the present is a necessary consequence of the past, and this implies that our present actions are the necessary consequence of the past. Fischer says "I do in fact find the Consequence Argument highly plausible, and I am inclined to accept its soundness. I thus think it is important to argue that there is an attractive kind of compatibilism that is indeed consistent with accepting the Consequence Argument as sound."¹⁵⁶ Semicompatibilism is precisely this kind of compatibilism. The account focuses on the compatibility of moral responsibility and determinism. Since a lot of the concern in free will debates over the freedom to do otherwise is related to moral responsibility, semicompatibilism pries apart the ability to do otherwise from moral responsibility and solely focuses on the latter.

¹⁵⁵ Peter Van Inwagen, *An Essay on Free Will*. (Oxford: Clarendon Press 1983).

¹⁵⁶ John Martin Fischer, "Compatibilism." In *Four Views on Free Will*, First Edition. (Malden, Massachusetts: Blackwell Publishing, 2007): 44-84.

Alongside Mark Ravizza, Fischer make a distinction between two kinds of control: regulative control and guidance control.¹⁵⁷ On the first model of control, an agent has control over the occurrence or nonoccurrence over an event as they are able to act and refrain from acting. On the second model of control, an agent exercises control of an event occurring or not occurring when they are able to make an event happen.¹⁵⁸ The difference in these models is best understood in terms of what each means by “control”. Guidance control entails that an agent has the relevant ability to bring about an event. Regulative control requires the same ability plus the ability to refrain from bringing about an event. Imagine that you are out driving, and you want to go to a restaurant, so you “guide your car to the right,” into the parking lot. You had “guidance control” at the time in those circumstances by virtue of guiding it through steering it to the right. Now suppose that there are no “special” factors at play—no evil demons or well-intentioned neurosurgeons installing microchips into your brain or any relevant disorder.¹⁵⁹ Then presumably you also had at the time you decided to turn into the restaurant the ability to refrain from going. These two models of control can come apart in the following way: imagine the car has developed a problem, so it will turn right in the next few seconds regardless of how the wheel is turned. If it’s turned right, it will operate normally, but if it’s turned left (or left alone), the car will also turn right. You turn the wheel to the right, guiding the car into the lot. You have guidance control but lack regulative control: no matter what you did, the car was going right.

Fischer argues that both models of control are possible but that guidance control is all that is needed for moral responsibility. Crucially, guidance control is compatible with determinism. The

¹⁵⁷ John Martin Fischer and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility* (Cambridge: Cambridge University Press 1998).

¹⁵⁸ Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*.

¹⁵⁹ Fisher and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*.

reason Fischer provides to support this is that we can control events without having control over them. Consider the following case: Jones has procrastinated about who to vote for in the upcoming election to the last possible moment. Jones deliberates in the voting booth in a “normal” way and chooses to vote for the Democratic candidate. What Jones does not know, however, is that a neurosurgeon (Black) has implanted a chip into Jones’ brain to make sure Jones votes Democrat. Black monitors Jones’ brain and if Jones votes Democrat she does nothing but monitor. If she sees that Jones is choosing to vote Republican, she swings into action and activates the chip to override Jones’ deliberation and make him vote Democrat. It just so happens that Jones votes Democrat on his own without any intervention by Black. Fischer argues that since Black and her chip play no role in his deliberation to choose to vote Democrat, Jones is morally responsible for his voting Democrat. Jones exercises guidance control for voting Democrat but lacks regulative control over his voting because he had no control over casting a ballot for the Democratic candidate instead of the Republican; he did have guidance control of his casting of the ballot. He has guidance control in the same way one has control of turning the car to the right. Jones deliberates over who to vote for and even though he is unable to vote Republican, his deliberation terminates all on its own in choosing to vote Democrat. He has control for voting because he deliberately chooses Democrat and he picks up the ballot and marks Democrat. Therefore, regulative control is not needed to assign moral responsibility to agents for their actions.

It is interesting to point out that this example has overlapping characteristics to Locke’s earlier case of a man trapped in a room. Jones and Locke’s man both do what they want to do but they could not do otherwise than they did. It seems that even in semicompatibilism, there is still a belief that doing what one wants to do is the relevant criterion. Of course, it would be a

misunderstanding to say they are the same, as semicompatibilism offers more than just an account of moral responsibility grounded on an account of agent's motivations at the time of action.

Consider the following variation of Fischer's car scenario: imagine that an individual I will call Harry is out driving on his way to work. Harry is perpetually late for work as he always misses the correct turn. Harry's friend, Jacob, being concerned about Harry's tardiness, has rigged Harry's car to turn right at the correct exit. Harry is blissfully unaware that Jacob has tinkered with his car. The usual exit comes up and Harry concentrates on not being late yet again. Since he does not know that the car is going to make the turn regardless of what he does, he believes he could turn left when he needs to turn right. His motivation to be on time results in his turning the wheel and guiding the car right. Harry lacked regulative control because he lacked control over the car turning right or left, but he did have control of the car because he steered it to the right. Harry makes it on time to his job; the lesson the semicompatibilist would draw is Harry deserves to be praised and hence that praiseworthiness only requires guidance control.

Fischer and Ravizza articulate a novel theory of moral responsibility because of the separations of moral responsibility from freedom. Fischer and Ravizza's view is unique in positing that agents can be morally responsible and lack freedom understood as the freedom to otherwise. As I outlined earlier in this chapter, other compatibilists believe that the freedom to do otherwise is possible if determinism is true. They merely added qualifications that the freedom to do otherwise is a hypothetical ability as opposed to a categorical one. As we already saw, Fischer and Ravizza's semicompatibilism concedes to the incompatibilism that one cannot do otherwise if determinism is true, but that we are still moral responsible. As long as we have guidance control for our actions, we can still be morally responsible. The obvious route for critical

commentary is that the semi-compatibilist believes alternative possibilities are irrelevant to moral responsibility. I think though a different approach would be better and this I must save until later on in the chapter as it is not connected to the next position I will now consider.

d) *Non-Reflectivism*

One last compatibilism to consider in this section is the view put forth by John Doris in his book *Talking to Ourselves: Reflection, Ignorance, and Agency*.¹⁶⁰ Doris argues that the doctrine he calls “reflectivism” is false because it does not line up with relevant empirical work in psychology. Reflectivism is the view that “the exercise of human agency consists in judgement and behavior ordered by self-conscious reflection about what to think and do.”¹⁶¹ He advances a theory of agency that is anti-reflectivist—it does not require self-conscious reflection to exercise agency—and it is also valuational—it locates agency in the expression of one’s values.¹⁶² To aid in our understanding, it would be good to get clear on what is meant by “agency”. A straightforward sense of agency involves just the performance of action; one is an agent just in the case that they do something. The sense of agency Doris takes aim at, reflectivism, involves more than just the mere performance of action.¹⁶³ Reflectivism is a kind of agency involving the conscious reflection and execution of conscious thought about what to do and think.¹⁶⁴ The anti-

¹⁶⁰ John Doris, *Talking to Ourselves: Reflection, Ignorance, and Agency* (New York: Oxford University Press, 2018). Utilizing cases from social psychology Doris deduces that reflectivism is an empirically false perspective of human agency. The cases are interesting but they would not advance our understanding of the view and might sidetrack the discussion in an undesirable way so I leave them aside.

¹⁶¹ Doris, *Talking to Ourselves*, 2.

¹⁶² Doris, *Talking to Ourselves*, 3.

¹⁶³ Doris, *Talking to Ourselves*, 3.

¹⁶⁴ Doris, *Talking to Ourselves*, 4.

reflectivist view of agency Doris advocates believes agency is the ability to act in such a way as to express our values when we act. As we will see shortly, Doris claims that our agency involves values over time and this makes his preferred sense of agency diachronic.

Consider Huck Finn from Mark Twain's *The Adventures of Huckleberry Finn*. Doris offers two interpretations of a critical moment in the novel where Huck faces a moral dilemma. He is internally torn between handing his friend Jim (an escaped slave) over to the authorities and aiding Jim in escaping to freedom. On the first reading, Huck helps Jim get away because of self-ignorance; Huck believes himself to hold the typical values of ante-bellum white southern Americans. However, it is clear he values his friendship with Jim more when he fails to report Jim to the authorities at the moment when "rubber hits the road." On the second reading, Huck is not ignorant of his values but he actually does value turning Jim in *and* helping him escape capture.¹⁶⁵

Doris describes Huck on this second reading as being ambivalent about what to do¹⁶⁶. Doris describes this ambivalence in relation to Frankfurt's views on what it means to be ambivalent. Frankfurt is well-known for saying that ambivalence is antithetical to free will because it undermines one's ability to be wholehearted about one's desires. The way Frankfurt describes ambivalence is as "a disease of the will" in the sense of lacking any preference for one particular desire and a corresponding course of action to be the one an agent wants and performs. In other words, there is no inclination for a particular desire to be the desire that motivates one to act.¹⁶⁷

¹⁶⁵ Doris, *Talking to Ourselves*, 160.

¹⁶⁶ Doris, *Talking to Ourselves*, 160.

¹⁶⁷ Harry Frankfurt, "The Faintest Passion." In *Necessity, Volition, and Love* (Cambridge: Cambridge University Press, 1999): 9-10.

To quote Frankfurt directly, the ambivalent agent “is inclined in one direction, and he is inclined in a contrary direction as well; and his attitude towards these inclinations is unsettled. Thus, it is true of him neither that he prefers one of his alternatives, nor that he prefers the other, nor that he likes them equally.”¹⁶⁸ In the same paper Frankfurt also says, “To the extent that a person is ambivalent, he does not know what he really wants.”¹⁶⁹ Ambivalence is an obstacle to agential unity which, Frankfurt thinks, is a necessary condition for free will and moral responsibility.

Doris rejects this conception of ambivalence and its antagonistic relationship to agency. Doris argues that “People aren’t all in all that much, it seems to me. That’s life... While the disintegration of mind... may destabilize agency... the exercise of agency does not require ‘unity’ understood as the absence of conflict... No doubt, the multiplicity of values with which people live sometimes make it difficult to say where their treasure lies, but that doesn’t mean there’s no hope of uncovering it, and still less that there’s no treasure.”¹⁷⁰ He also believes “... much human life occupies middle ground, where all a person’s values don’t speak with one voice.”¹⁷¹ Contra Frankfurt, wholeheartedness is not a necessary condition for free will and ambivalence is not an obstacle to it. As Doris sees it, ambivalence is a part of life itself because human agents contain a multiplicity in regard to what they value, and it follows that we will be repeatedly ambivalent throughout our lives. There will be moments where we struggle to act but it does not follow that, because we were not decisive throughout the deliberative process, we did not act as agents. Doris seems to be in agreement with Kane that being ambivalent and resolving

¹⁶⁸ Frankfurt, “The Faintest Passion,” 10.

¹⁶⁹ Frankfurt, “The Faintest Passion,” 9.

¹⁷⁰ Doris, *Talking to Ourselves*, 162.

¹⁷¹ Doris, *Talking to Ourselves*, 161.

it through action is a distinctive feature of agency. On the second reading of Huck, whatever he decided to do the act counts as an expression of agency at the time in virtue of that action expressing his values. Ambivalence was not antithetical to Huck's agency but actually a part of it. It might have been difficult and Huck might have had to struggle to choose because he was ambivalent. The nature of the ambivalence is rooted in Huck's internal valuational conflict.

Using the distinction between freedom of action and freedom of will, I think we can see that Doris has an important point to make about agency and moral responsibility as Doris views the two as related.¹⁷² As a quick refresher, freedom of action is a category of freedom that encompasses a range of abilities that are all grounded on the ability to do what one wants to do. Freedom of will is the ability to shape one's will the way one wants. It is possible for one to have freedom of action but not freedom of the will as we saw in the accounts from Watson, Wolf and Frankfurt. Doris' account fits into this distinction in a couple of different ways. Doris, like the authors just mentioned, offers a way of understanding how one could be said to act freely but not have free will. One possibility is that we are so wrong about our motivations when we act, that on one level we act freely, since we do what we want, but because these motivations are not authentic they are not really our own, and so our wills are not truly free. Another related feature of Doris' view is his conception of ambivalence because it is alleged to be distinctive of agency. This ties into the distinction under discussion because in his view this shows value, and therefore agency, has a diachronic dimension.¹⁷³ Doris' theory therefore believes that agency is about more than just doing what one wants but about one's will over time.

¹⁷² Doris, *Talking to Ourselves*, 165.

¹⁷³ Doris, *Talking to Ourselves*, 163.

Doris' understanding of the distinction between action and will is clearly seen when he describes this feature of agency which follows from his view of value and ambivalence. He writes

The exercise of agency emerges not only in discrete, synchronic, actions, but also in temporally extended processes. In many endeavors—a relationship, career, or cause—the exercise of agency is temporally developed, which means that attributions of responsibility may attach to processes as much as actions. This realization is not especially evident in philosophical writing on agency and responsibility, which characteristically focuses on more or less discrete behavioral units... rather than complex temporally extended processes. The second query wants explication in terms of values... while the first sort of question may often be answered without direct reference to values. Of course, extended performances don't always require valuational explanations; But the answer you get depends on the question you ask, and by asking about ongoing processes rather than discrete actions, you may end up with different sorts of answers.¹⁷⁴

By thinking of agency in terms of extended processes, a feature of free will can be brought out that I think is significant: free will has a diachronic nature to it in the sense that one can freely choose not just a discrete action but also temporally extended actions. To choose in a moment what to have for supper is a discrete action, while deciding what career to choose is a temporally extended one. For instance, if I choose to become a lawyer, I have to choose which law school to go to, which area of law to specialize in, and I have to finish law school and start practicing before I can say I am a lawyer. The sense of free will that I am interested in sometimes involves this kind of temporally extended processes. I say "sometimes" to include both extended processes and discrete units. Free will as I understand it involves, extended processes through the choice to continue or to not continue an activity. Instead of a self-forming action, consider self-forming extended processes. Choosing to be a lawyer is partly a discrete event as it involves individual actions like accepting an admission offer, but it is a temporally extended process with different possible directions. One could go to the local law school and become a lawyer

¹⁷⁴ Doris, *Talking to Ourselves*, 163.

specializing as a defence attorney or move away to an ivy league school and become a corporate lawyer. Even after one has chosen, one could face moments of hesitation and be tempted to abandon the project chosen and try to go the other direction.

We can derive another feature of free will from this line of thought and this is that through choosing whatever they do agents can use their free will as a tool of self-knowledge and self-shaping. I think that Doris' discussion of Huck is open to this sort of interpretation. Huck is ambivalent about his values and, in keeping with Doris' view of ambivalence, it seems plausible to say that when an agent chooses in such a moment of ambivalence, they could come to learn about themselves. An agent who constantly prioritizes one project over another across a temporally extended period of time can, though need not, come to learn something about themselves through their actions. More importantly, the free choices which shape us are an important part of the sense of personal identity that I am interested in defending. This view of identity explains personal identity in terms of personal characteristics which endure over time. This sense of identity seems to align with the current diachronic aspect of agency and free will as the self-forming choices made in the context of an extended process help to explain the connection between identity and free will as I understand it.

3. Internalism and Externalism

Let's classify compatibilism into two kinds: internalist and externalist. Broadly speaking, an account of free will and moral responsibility is internalist just in case the view only considers an agent's internal psychological properties at a time. An account of free will and moral responsibility is externalist just in case the view also considers external features of an agent's

psychological economy, broadly speaking, as they pertain to the production of action.¹⁷⁵

Externalists acknowledge that:

the freedom-and-responsibility-conferring psychological structure sought at a moment in time is partially determined by the causal history giving rise to that structure at that moment in time. That history will include causal features stretching outside the agent herself, such as the manner of education and socialization that shapes a person's values and beliefs.

Externalist theories are thus *historical theories* (author's italics).¹⁷⁶

Let's begin with the following case: Margie is deliberating over whether to cheat on her taxes.

She reflects on whether this would be consistent with her values. She concludes, rightly, that it would be consistent with her values to cheat on her taxes and does so. We can raise the question of Margie's moral responsibility and freedom of will from both an internalist and externalist perspective. On an internalist reading of the case, Margie is free if whatever internal property/properties one takes to be necessary and sufficient for moral responsibility and free will were present and exercised. The causal origins of these properties are not relevant. So, if we said that acting in accordance with your values was what clinches the judgement we ought to make about Margie's freedom and moral responsibility, all we need to consider is if she did indeed act in alignment with her values. We have stipulated that she did and so we should conclude she freely cheated on her taxes and deserves to be blamed for doing so. The externalist reading of the case would say the properties of an agent cannot be considered in isolation. If we say that the property needed to act freely and be blameworthy is acting according to our values, we have to consider the way those values might have been shaped by influences external to the agent, such as her upbringing, education and socialization. An externalist would say Margie cheating on her

¹⁷⁵ Coates, Justin D. and Michael McKenna. "Compatibilism", *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), Edward N. Zalta (ed.).

¹⁷⁶ Ibid. In a technical sense externalist theories could be nonhistorical as one could restrict the scope to the immediate external environment. The most important external phenomena for the present topic though are historical and I am going to be restricted in my discussion to historical theories.

taxes might be consistent with her values, but we have to at least consider the circumstances that helped influence those values as they have the potential to undermine her freedom and moral responsibility.

Mele helpfully distinguishes varieties of internalism and externalism. He offers the following four definitions:

1. Unconditional internalism: An agent's internal condition at the relevant time and the involvement of that condition in his *A-ing* are relevant to whether he is morally responsible for *A*, and no fact about how he came to be in that condition is relevant.
2. Conditional internalism: There is some internal condition *C* such that if an agent who purposefully *A-s* is in *C* at the relevant time and *C* (or a part of *C*) accounts for *A* in internal way *W*, the agent is morally responsible for *A* no matter how he came to be in *C*.
3. Conditional externalism: Even if some conditional thesis is true, agents sometimes are morally responsible for *A* partly because of how they came to be in the internal condition that issues in their *A-ing*; and, more specifically, in these cases, there is another possible way of having come to be in that internal condition such that if they had come to be in that condition in that way, then, holding everything else fixed (except what is entailed by the difference, of course), including the fact that they *A-ed*, they would not have been morally responsible for *A*.
4. Unconditional externalism: Conditional internalism is false.¹⁷⁷

We have seen an example of unconditional internalism with classical compatibilism. The condition, on their view, is having a desire which motivates you to act. How you came to have that desire is irrelevant for those theorists. Frankfurt and Watson offer examples of conditional internalism. Their accounts propose a candidate for *C* and no fact about how they came to be in *C* is relevant to their moral responsibility. Wolf is interesting because of the asymmetry in her view between praiseworthiness and blameworthiness. Wolf's view counts as conditional internalism in the case of praiseworthiness but she belongs in the conditional externalism camp when it comes to blameworthiness. It matters how we come to be in *C* in case of assigning blame

¹⁷⁷ Alfred Mele, *Manipulated Agents: A Window to Moral Responsibility*, (New York: Oxford University Press, 2019): 13-14.

but not praise. How JoJo came to lack *C* matters for blaming him. Fischer is also another interesting case because work earlier on in his career would seem to make him a conditional internalist but the work done with Mark Ravizza does consider history in a way that makes it count as conditional externalism.¹⁷⁸ Mele himself has offered a kind of compatibilism that is intended to be a conditional externalism.¹⁷⁹ I am inclined to place Doris' account in the conditional internalism camp, though the way he tells Huck's story I would not resist putting him in the conditional externalist category. On Doris' account, moral responsibility comes down to the expression of one's values and it is acceptable that an agent is determined to hold the values they have.¹⁸⁰ The way he talks about Huck choosing his values, however, brings history into our considerations about moral responsibility and free will.

With this brief categorization of compatibilism in place we can get to what is an interesting divide among compatibilists and between compatibilism and incompatibilism: the significance of history in free will and moral responsibility. Mele's formulation of internalism and externalism helpfully sharpens this divide. Compatibilism that takes the form of either unconditional or conditional internalism is an ahistorical view.

Mele has a case he uses to show that 1 is false. Van and Ike both get drunk and kill a pedestrian while driving. Van willingly overconsumed alcohol, knew the effects of what this does, and willingly got in the car. Ike was forced to drink and forced into the car. Mele says Van

¹⁷⁸ Fischer and Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*.

¹⁷⁹ Mele, *Manipulated Agents*.

¹⁸⁰ Doris, *Talking To Ourselves*.

is blameworthy and Ike is not, and that the reason why is due to their history. Ike was forced to do it and Van was not.

Both Frankfurt and Watson are explicit in denying history as being relevant for judgements of either free will or moral responsibility. Frankfurt says:

A manipulator may succeed, through his interventions, in providing a person not merely with particular feelings and thoughts but with a new character. That person is then morally responsible for the choices and the conduct to which having this character leads. We are inevitably fashioned and sustained, after all, by circumstances over which we have no control. The causes to which we are subject may also change us radically, without thereby bringing it about that we are not morally responsible agents. It is irrelevant whether those causes are operating by virtue of the natural forces that shape our environment or whether they operate through the deliberately manipulative designs of other human agents.¹⁸¹

In a similar spirit Watson writes:

For the compatibilist, the constitutive conditions of free agency do not conceptually depend on their origins. In this sense, free and responsible agency is not an historical notion. Consequently, compatibilism is committed to the conceptual possibility that free and responsible agents, and free and responsible exercises of their agency, are products of super-powerful designers. For consider any compatibilist account of the conditions of free agency, C. It is possible for C to obtain in a causally deterministic world. If that is possible, then it is possible that a super-powerful being intentionally creates a C-world, by bringing about the relevant antecedent conditions in accordance with the relevant laws. This possibility follows from the general point that the conditions of responsibility do not necessarily depend upon their causal origins.¹⁸²

Watson adds: “If we define a robot as a creature whose existence and detailed program were brought about by design, then the foregoing reasoning commits compatibilists to the possibility that free agents are robots.”¹⁸³ These remarks suffice to establish the internalist as committed to the belief in the irrelevance of history to free will and moral responsibility. We should tread

¹⁸¹ Harry Frankfurt, “Reply to John Martin Fischer.” in *Contours of Agency: Essays on themes from Harry Frankfurt*. eds Sarah Buss and Lee Overton, (Cambridge, Massachusetts: MIT Press, 2002): 28.

¹⁸² Garry Watson, “Soft Libertarianism and Hard Compatibilism,” *Journal of Ethics* 3 (1999): 360-361.

¹⁸³ Watson “Soft Libertarianism and Hard Compatibilism,” 361.

carefully here but this commitment gives us good reason to be skeptical of this sort of compatibilism. Now I am not arguing that we should reject the possibility that certain histories are undermining of free will while other are not. What I am suggesting is that the line of reasoning present in Watson rejects history as relevant in any way and this is what should be scrutinized. We are part of a causal environment and to say that the way this has influenced certain aspects of us like our values is not relevant suggests a picture of human agents as independent from that environment. It paints a picture of agency resting in what Berlin called the “inner citadel” where our freedom is untouched by any of the causal influences which shape us.¹⁸⁴ A more accurate picture of agency does not just acknowledge the influence of history, but contends it is the present state of mind in judging moral desert.

Consider the following stories which are inspired by Mele:

Evil Edgar: Edgar is a person with values that judge murder as something that is the best thing to do. When he murders it is because it is what he wants to do. He considers his bloodthirsty desire to be an integral part of himself and wholeheartedly endorses it. When Edgar was younger he enjoyed torturing and murdering animals. But he was not wholeheartedly behind this desire because he felt some guilt over his deeds. However, Edgar valued being the kind of person who does what he wants and views moral systems which condemn murder as designed by the weak to chain the strong. Edgar hardens his heart to ignore any feeling of guilt and indeed he becomes thoroughly desensitized to murdering animals and eventually people. Edgar in pursuit of his goal sought out victims he could brutalize and he did this with increasing frequency. He eventually moves on from animals and kills persons for the sheer satisfaction of killing. He seeks out a victim, Chris, who Edgar believes he can kill and no one will notice his disappearance and so get away with his crime. He stabs him in the heart and disposes of the body. Howard is not insane because he recognizes that conventional morality says what he has done is wrong but he just does not care.

Saintly Susan: Susan is one of the nicest people of the world. Her values are those of someone who would never harm a fly. Susan volunteers in the local community and helps the Girl Scouts with their fundraising events. When Sally was younger, she reflected on herself one day and did not like what she saw. She thought she was too self-centred, petty and only did things she thought would directly benefit herself. She worked hard at changing herself and succeeded. She seems to be partly morally responsible for her

¹⁸⁴ Isaiah Berlin, *Liberty*, eds Henry Hardy, Oxford: Oxford University Press 2002: 120.

values in light of the fact that she consciously chooses to pursue them and succeeded in doing what she wanted to do. One day she falls asleep and awakes with a desire to stalk and murder her neighbor James. She is very disturbed by this as this desire is at odds with the other values she worked hard to cultivate. What happened is that while she was asleep a covert team of scientists discovered how to transplant Edgar's values into Susan and memories intact which is why she is surprised by this desire. Sally reflects on this new desire. While initially disturbed her reflection makes her see that this desire is aligned with her value system. She now believes that morality is a lie created by the weak to hold back the strong. Reflecting further on this desire she wholeheartedly endorses it and it is an expression of her values. She stalks and kills James making sure she does so in a fashion where she will not get caught.

Mele says that these two agents are value twins in the sense that the contents of their values are the same.¹⁸⁵ They meet all of the conditions that the conditional internalist says are needed for moral responsibility. According to Mele, Sally is not morally responsible for her action in the same way Edgar is. Sally's action was determined by a causal history (not necessarily itself deterministic or indeterministic) that involved manipulation, while Howard chooses to become a bad person. Sally and Howard both meet the conditions stipulated to be sufficient for moral responsibility and free will, but this is a problem for such an account. It is a problem because the introduction of historical considerations results in different judgements about moral responsibility and free will. This is contrary to Frankfurt's and Watson's assertion that these concepts are ahistorical. Even though the outcome in both cases is the same the process is not irrelevant in the cases when assigning moral responsibility.

It is interesting that Mele believes that if it rested on Frankfurt's version of internalism, compatibilism would be in dire straits.¹⁸⁶ What is intriguing about this claim is Mele's conviction that compatibilism needs some kind of historical consideration included in its analysis of moral responsibility and free will. He thinks that, "any competent compatibilist will claim that *some but*

¹⁸⁵ Mele, *Manipulated Agents*, 21.

¹⁸⁶ Mele, *Manipulated Agents*, 22.

not all deterministic causal routes to action preclude moral responsibility” (authors italics).¹⁸⁷ If we take a look at the history involved in Susan and Edgar, we will make different judgments about their responsibility. The reason their histories generate such a difference in responsibility is because of Susan’s values not being something she helped to cultivate. Imagine that Susan was abducted by scientists and they told her they were going to implant Edgar’s values into her mind. The scientists will wipe her mind before performing this implantation and at that moment, Susan would be appalled because she had tried so hard to be a good person. If she would reject the values the scientists implanted in her. Edgar, on the other hand, never once flinches from the person he is and becomes and, indeed, does everything the can to become who he is. We might say that Susan should be said to have shared responsibility and not less. It is perfectly true that people can individually be assigned blame for a shared action. But this misses the point in my view of the case, which is that, through taking into account her history, we can see Susan involuntarily becomes a different person and her values are not voluntarily formed. She is deterministically controlled to act differently than she would have wanted to without the intervention of the controllers.

Mele believes compatibilism should make conceptual space for history and advance a conditional externalist approach.¹⁸⁸ An important motivation for Mele in advancing a historical sensitive compatibilism is trying to answer incompatibilist concerns about history and moral responsibility. He is specifically attempting to construct the resources needed to reply to a

¹⁸⁷ Mele, *Manipulated Agents*, 28.

¹⁸⁸ Mele, *Manipulated Agents*, 32.

concern that Kane has over what he calls “CNC control.”¹⁸⁹¹⁹⁰ CNC control is comprised of two distinct kinds of control but can be put together: it is made up of covert control and non-constraining control. Covert control is the control of an agent by another agent without the controlled agent being aware of the control. It is possible that the covert nature of the control be carried out with the controller’s existence being known. Non-constraining control is controlling an agent without using coercion. We can see CNC control at work in the Susan case, and it does a lot of the work in Mele’s argument regarding the proper judgement of moral responsibility. Mele wants to show how being determined to act by the control of other agents is different from being determined by naturalistic forces. Incorporating history in the way that the Edgar and Susan cases do will provide the compatibilist what she needs to handle concerns about history.¹⁹¹ I think that a history-sensitive compatibilism is the right way to go if one were to advance compatibilism over incompatibilism.

4. Historical Compatibilism or Standard Incompatibilism

With a more nuanced compatibilist account on hand from Mele, we can consider it alongside the incompatibilist view I articulated in the previous chapter. A history sensitive compatibilism is a welcome addition to discussions about freedom of the will and moral responsibility. Some of the intuitions that inspire incompatibilism come from worries about personal history. The examples of Robert Harris and the young man Kane talked about are prime examples of concerns about the past influencing us and potentially mitigating our moral responsibility for our actions. The externalist perspective discussed shows how this could be true

¹⁸⁹ Kane, *The Significance of Free Will*, 72.

¹⁹⁰ Mele, *Manipulated Agents*, 1.

¹⁹¹ Mele, *Manipulated Agents*, 2.

because both of these cases involved causal histories with influences that were so great as to leave nothing open to the men to control. Their upbringing was so abusive and damaging that the values and desires which motivated their behavior did not come from anything they chose or decided to do but were shaped by external forces that made them what they were.

Incompatibilists like myself believe these agential historical considerations start one down the path to the incompatibility of determinism with responsibility. Mele's historical compatibilism would seem to offer a way for the compatibilist to reply here by saying that the influences which shaped Harris and the young man are different in kind from mere natural forces. The deterministic causal histories involved in these cases are different from deterministic causal histories which do not involve abusive upbringings or poor circumstances. If Robert Harris was raised in a loving home and still grew up to be a cold-hearted killer, then even if determinism were true Harris is morally responsible for his actions. The compatibilist has the resources with Mele's account to separate deterministic causal histories that incompatibilists would say show responsibility is incompatible with determinism from deterministic histories which do not threaten moral responsibility.

Mele's historical sensitive compatibilism distinguishes between being deterministically controlled and being the product of deterministic natural processes. This is a move, as we will see in chapter four, that Dennett makes and the way it works is to say that agential control determining agents is to kept separate from non agential control determining agents. I think this is harder to do than it might appear at the surface. The introduction of the argument needed to collapse this distinction would take the present chapter on an undesirable detour so instead we now turn directly to the issue of incompatibilism.

Chapter 4

Incompatibilism

1. Introduction

In this chapter, I attempt to motivate an idea that finds its roots in the soil of incompatibilism: ultimate sourcehood. Kane brought this idea into prominence, and I follow Kane in postulating ultimate sourcehood to be a necessary condition for free will.¹⁹² In my view of free will, a plausible account of ultimate sourcehood involves models of control that are friendly to compatibilism but expands on these models to include incompatibilist control; incompatibilist control develops out of compatibilist control. Compatibilism offers essential pieces of the free will puzzle, and it is a mistake to deny this. Requiring ultimate sourcehood is consistent with compatibilism since such things as guidance control, sanity, acting by one's will, etc., are all necessary conditions for ultimate sourcehood. This chapter focuses on the ultimacy component of ultimate sourcehood. The discussion motivating ultimacy starts with discussing Frankfurt's theory of free will and a prominent objection to his view. Frankfurt's theory in essence says that free will boils down to having the will that one wants to have. The objection I explore focuses on the origins of an agent's will and casts doubt on the sufficiency of Frankfurt's criterion of wholeheartedness for free will. Using this objection as fodder, I introduce the motivation for thinking that an agent must be the right source for her will for an agent's will to be free. I argue that this requires an agent be the ultimate source of her will and offer an argument regarding the nature of ultimacy. I lastly offer a further refinement of the position through engaging with John Martin Fischer's work on ultimate sourcehood. The roadmap for the chapter is the following: in the next section I introduce the objection to Frankfurt's view and use it to build a bridge to

¹⁹² Kane, *Significance of Free Will*.

sourcehood. I discuss concerns around an agent's will and the satisfaction of the conditions for one's will to be free. The third and fourth sections introduce and discuss features of ultimacy while the fifth and sixth sections offer a defense and refinement of ultimate sourcehood.

2. Hierarchical Theories

Let's briefly summarize Frankfurt's view from chapter three and then turn to a prominent objection raised against it, which will then bring us to the topic of this section. Frankfurt begins his account of freedom of the will by first observing what makes human persons different from non-human animals. He says that both animals and persons have desires, beliefs, and perform actions, but only humans have "second-order desires" or "desires of the second order."¹⁹³ Second-order desires are desires about first-order desires. Consider an agent who has a first-order desire for cheeseburgers. His first-order desire is simple to understand as it is just the desire or want for cheeseburgers. His second-order desire is not to want cheeseburgers. Frankfurt's account adds another important component that brings self-reflection into the picture in the form of his technical notion of a 'volition.' This plays a demarcating role in his view as it separates persons from what Frankfurt labels as a 'wanton'.¹⁹⁴ Persons possess volitions while wantons do not; persons care about which desires they have motivate them to act while a wanton merely acts for whatever desire moves them accordingly. Frankfurt's account is an improvement over classical compatibilists like Thomas Hobbes who believed that all that mattered to free action was the absence of impediments to doing what one wants to do. Frankfurt's view allows us to see free will is not quite the same as free action though they are related.

¹⁹³ Frankfurt, *The Importance of What We Care About*, 12.

¹⁹⁴ Frankfurt, *The Importance of What We Care About*, 17.

A persistent critique of Frankfurt, first developed by Richard Double, Dean Zimmerman, David Shatz, and Gary Watson, takes aim at the nature of second-order volitions.¹⁹⁵ For example, Double asks, “Suppose I decisively identify with my first-order desire... to sacrifice my life if my religious leader asks me to. If I do so in a completely nonreflective way, it is difficult to see how I have more freedom than a wanton, despite the fact that according to Frankfurt’s ... account these decisions are mine.”¹⁹⁶ This critique presents Frankfurt’s view with a dilemma: Double’s point is seemingly that unreflective identification, especially prompted by others, would seem not to establish agents as free; reflective identification seems like the solution to this point but there is a serious problem. Consider what it would mean for me to reflect on my desires: when I reflect on my first-order desires I think about whether I really want to want or want not to want my first-order desires, but these second-order desires now seem unreflective, so reflection on my second-order desires is needed, which brings in unreflective third order desires, and so on. Attentive readers will likely have thought that this higher-order reflection seems to have no end. Frankfurt’s account seems committed to an infinite regress of reflection once we start reflecting on our first-order desires. So, either way we spin the view we do not seem to have the sort of freedom Frankfurt desires.

Frankfurt’s solution to this problem is to appeal to technical notions of *identification* and *wholeheartedness*. An agent identifies with their second-order desires when “there is no room for questions concerning the pertinence of desires or volitions of higher-orders.”¹⁹⁷ Identification

¹⁹⁵ Gary Watson, “Free Action and Free Will,” *Mind*, 96, No. 382, (1987): 152; David Zimmerman, “Hierarchical Motivation and the Freedom of the Will,” *Pacific Philosophical Quarterly* (1981): 354-368.; David Shatz, “Free Will and the Structure of Motivation,” *Midwest Studies in Philosophy* (1985): 451-82.; Richard Double, *The Non-Reality of Free Will* (Oxford: Oxford University Press, 1991).

¹⁹⁶ Richard Double, *The Non-reality of Free Will* (Oxford: Oxford University Press, 1991): 35.

¹⁹⁷ Frankfurt, *The Importance of What We Care About*, 11.

alone does not satisfy Frankfurt's critics, such as Watson, who writes, "We wanted to know what prevents wantonness about one's higher-order volitions. What gives these volitions any special relation to 'oneself'? It is unhelpful to answer that one makes a decisive commitment where this means that an interminable ascent to higher orders will not be permitted. This *is* arbitrary."¹⁹⁸ Frankfurt responds to this by articulating the requirements of wholeheartedness. If one is wholehearted about the desires one identifies with, one is fully satisfied and has no interest in reflecting on higher-order desires. He says, "It is a matter of simply *having no interest* in making changes."¹⁹⁹

Frankfurt's rejoinder is not entirely satisfactory to such critics as Watson who brings in a concern regarding the origin of higher-order preferences. Watson summarizes the worry precisely, writing, "For all [Frankfurt's] account tells us, the person's higher-order preference may be the result of brainwashing, or severe conditioning of the kind of which is incompatible with autonomy."²⁰⁰ Watson says "autonomy" is incompatible with brainwashing; he uses "autonomy" interchangeably with "free will." Watson implicitly says here that a preference or presumably any psychological state that is the result of brainwashing or severe conditioning is incompatible with autonomy or free will. Watson is not satisfied with Frankfurt's proposal that as long as we are wholehearted about our desires than we have free will. Here is how I interpret his dissatisfaction: Watson's worry with Frankfurt's account stems from Watson's concerns about the origins of the things that satisfy the conditions of the freedom of the will. In Frankfurt's view, the alignment of our higher-order and first-order desires culminating in action

¹⁹⁸ Gary Watson, "Free Agency," *Journal of Philosophy*, 72, no. 8 (1975): 217.

¹⁹⁹ Harry Frankfurt, "The Faintest Passion," 105.

²⁰⁰ Watson, "Free Action and Free Will," 148.

is what is needed for freedom of the will. If an agent's action matches what he really wants to do, his action and his will count as free. Consider the following case: James is an agent that meets Frankfurt's criteria for freedom of the will. James has a desire to do X and wishes for this desire to be effective in the production of action. He wholeheartedly endorses and identifies with the desire to do X. The alignment between James' desire to do X and his desire for X to be effective in action in conjunction with his wholehearted endorsement of X are the conditions for freedom of the will; James satisfies these conditions. Now, suppose that the satisfaction of these conditions has been produced by James being brainwashed. I will put some content into X to construct a case. Suppose Jacob has brainwashed James to desire and wholeheartedly endorse the desire to eat entire bags of potato chips. James satisfies Frankfurt's criteria for freedom of the will, but the satisfaction has been produced by brainwashing. Watson raises his complaint with the purported conditions of freedom of the will at this stage.

Let's distinguish between the manipulation of the will and the manipulation of the things purportedly needed to satisfy the conditions for freedom of the will. One could have their will implanted while satisfying the conditions of freedom or one could have their will naturally develop and the conditions for their will to be free be met through manipulation. Consider Frankfurt's account: there is the will itself and then there is wholeheartedly identifying with one's will. The criticisms surveyed in the present discussion present a challenge for an account such as Frankfurt's that proposes conditions to be met for freedom of the will that can be obtained through processes like brainwashing. To speak more generally, the issue is that an account of freedom of the will which has conditions that can be met via unendorsed implantation or manipulation through such things as brainwashing does not deliver freedom.

The distinction introduced between the will and satisfying the conditions for freedom of the will leaves an open question regarding an agent being the ultimate source of her will and not of conditions of her will being free. Frankfurt's theory describes the will as separate from satisfying the conditions needed for the will to be free. We can ask, if satisfying conditions like those offered by Frankfurt will not make an agent the source of her will as they are vulnerable to manipulation worries, then what will satisfy the conditions of freedom of the will? I see being the source of the will and the source of the satisfaction of the conditions of the will's freedom as related, and so manipulation of the conditions of freedom of the will is related to being the source of one's will. I argue that, to be the source of the satisfaction of the conditions of the freedom of the will, an agent must be the right sort of source of the will itself.

The right kind of sourcehood is ultimate sourcehood. Ultimate sourcehood is a nuanced concept and the defence of it will take place here and the next two sections. The first plank that has to be deployed and defended is that the conditions of freedom of the will include an agent as the ultimate source of her will itself. The nature of ultimacy is the next step that will be put forth and earned in the next two sections. What I aim to do here is earn the point that agents are an ultimate source of their wills through being the source of what satisfies the conditions of freedom of the will. I will consider cases where an agent meets certain conditions but in one case is not free because she is not the source of her will, while in the other case, she is free because she is the source of her will. These cases will powerfully suggest that to satisfy the conditions of freedom of the will, an agent must be the right sort of the will itself. On my view, being the source of one's will is both necessary and sufficient for the will to be free.

The sufficiency claim has a lower argumentative burden than the necessity claim. Take the businesswoman from chapter two: while on her way to an important meeting she witnesses a

person being mugged. She is free to do what she wants whatever she does, and she has reasons to support either course of action, and neither her reasons or her will itself are products of manipulation. Through something that she does the businesswoman is intuitively the source of her will as her choice effects the structure of her will. The outcome of her choice seems to come from her in that she is the source of the decision. Her choice is a cause that has the effect of shaping her will and as she is the source of her choice, she is the source of her will. Or consider Paolo, who for the sake of argument is stipulated to be the source of his will. Paolo decides to go on vacation and this decision comes from a desire internal to his will. It seems plausible that Paolo's choice to go on vacation is free because he is the source of his choice, as is the source of his will. Paolo's sourcehood regarding his will makes his choice free.

Establishing that it is necessary for Paolo to be the source of his will for it to be free is trickier than that it is to show that it is sufficient for his will to be free if he is the source of his will. I have to show that only agents who are the sources of their wills have free wills. What needs to be done is to work through cases where an agent is only the source of things that putatively satisfy the conditions of freedom. I will construct cases where an agent's will is manipulated but they satisfy a condition that is purported to deliver freedom. There are various conditions surveyed in chapter three. These conditions include sanity, responsiveness to reasons, acting in accordance with our wills, expressing our values, wholehearted identification and so forth.

The first condition of the freedom of the will comes from Watson. Watson's view of the conditions of freedom involves an agent having their valuational and motivational systems, values and desires, align with each other. Imagine Josh values taking care of his health and so he desires to eat whole grains, lots of fruits and vegetables, and to engage in sufficient exercise.

Sometimes Josh is tempted and gives in to his nonhealthy habits. When Josh acts in alignment with his values, he acts in accordance with his will, and when this is not the case he acts contrary to his will he exhibits weakness of will.²⁰¹ Imagine a slightly different case where Josh's will is the product of manipulation and he always acts in accordance with his will and so never acts due to weakness of will. Suppose that Josh's friend Jake is responsible for manipulating Josh's will so that Josh will never falter to temptation. Josh always chooses what he values and his desires never conflict with what he values. Jake makes Josh pick salad over french fries as a meal. Josh's motive and values align, but he does not choose the salad freely—Jake made it that Josh would choose it. Jake, not Josh, is clearly the source of Josh's will. As I argued in chapter three, acting from weakness of will is not always done unfreely. It seems quite plausible for Josh to be able to give in to temptation freely if we posit that he is the source of his will because the temptation comes from something of which he is the source.

The second condition that I will consider is sanity. As a quick refresher, the condition for free will proposed by Wolf was that an agent acts freely if they are sane while they are not free if they are insane.²⁰² Recall from the discussion in chapter three that Wolf raised her condition in the context of Frankfurt view of the conditions of freedom. Wolf constructs a case where an agent, JoJo, has been raised by his father who is a ruthless dictator and inflicts cruelty on his citizens. JoJo is socialized and taught that people who do not salute him correctly deserve to be tortured. JoJo comes to wholeheartedly identify with treating people inhumanely. However, it is quite clear that JoJo is insane because he cannot recognize his bad behaviour or respond to the moral reasons to not act this way. So, Wolf argues, the conditions of freedom Frankfurt are not

²⁰¹ Watson, "Free Action and Free Agency".

²⁰² Wolf, "Sanity and the Metaphysics of Responsibility," 185.

sufficient because JoJo satisfies them while clearly not being free because he is insane when he acts.

Consider a variation of the JoJo case: Suppose that JoJo's will has been covertly manipulated by Koko through brainwashing but that JoJo is sane. JoJo is not the source of his will, but he does satisfy Wolf's condition for freedom of the will. It is possible to show that despite JoJo satisfying Wolf's suggested condition for freedom of the will, not being the source of his will counts against his will being free. Suppose that JoJo makes a choice to punish one of his citizens. JoJo's action to do this is because of results from KoKo's covert manipulation. Even though JoJo satisfies the condition of sanity, JoJo does not have free will. JoJo is not the source of his will and he is not free even while satisfying the sanity condition because of KoKo's manipulation. Now suppose that JoJo was able to choose to punish and not punish the dissident and his will was not so manipulated that it entailed his decision, while also still being sane. When JoJo chooses, he now chooses freely because he is now the source of his will.

A third condition for freedom of the will explored earlier from Doris is the expression of our values. This condition is not focused on the alignment of our values and desires. Doris is more focused on simply expressing what we value when we act. Doris would be skeptical of an account like Watson's because it would fall neatly into a reflectivist view of agency which he argued was empirically false.

Now unlike with the discussion up to this point, I won't separate the will from the satisfaction of the conditions of freedom. This is because that as I understand it on Doris' view the separation is impossible: if an agent is the source of the things that satisfy the criteria for freedom, you are thereby necessarily the source of your own will. If an agent is the source of the

values they express, they are the source of satisfying the condition and necessarily the source of their own will. So if an agent is the source of their values they are the source of their wills.

Fischer's view also does not have a separation between the satisfaction of the conditions of freedom and the will itself. Fischer's view as I described is that agent's free will and moral responsibility is grounded in an agent's reason-responsiveness. An agent is able to satisfy this condition when they are able to act in response to reasons and they do not act out of compulsion. Imagine that Emily is deciding what to order for dinner on her smartphone. Emily does not know that the smartphone will only allow her to choose Italian. It turns out that Emily wants to order Italian cuisine for dinner and so she does. Emily satisfies this condition because there was reason to which she responds and the fact that she could not have done otherwise is irrelevant. If we imagine that it were possible for her to have done otherwise, had her reasons been different she would have responded differently. This case demonstrates how the satisfaction of this condition for freedom of the will operates. When an agent acts in response to reasons and they do not act compulsively they satisfy the condition for freedom of the will. For this condition of freedom to be satisfied, all that is required is they respond to a reason and that the agent is not acting from compulsion. When Emily orders food, her desire for Italian is the reason for what she does. Huck and Emily are cases that serve to demonstrate the sourcehood involved in free will: through being the source of satisfying the conditions of freedom Emily and Huck necessarily are the sources of their wills.

This case along with the other cases I have presented provide a powerfully suggestive argument for the necessity claim for sourcehood: the claim that agents must be the sources of the wills themselves to have free wills and this included being the source of satisfying the conditions for wills to be free. Through arguing that if an agent's will as well as the satisfaction of the

conditions of freedom are the products of covert manipulation the agent does not have a will that is free or freely satisfies the conditions of freedom. Combined with the earlier argument for the sufficient argument, I have put together a strong argument to say that sourcehood is sufficient and necessary for free will.

3. Ultimate Sourcehood and Responsibility

In the last section, I introduced the idea of ultimate sourcehood in grappling with manipulation. I now turn towards ultimate sourcehood directly. We can separate ultimacy from sourcehood as one can be a source without being an ultimate source. This is the strategy pursued by certain compatibilists attracted to this idea. Incompatibilists attracted to the idea of being a source arrive at incompatibilism with the addition of ultimacy. Ultimate sourcehood requires that to be ultimately free and morally responsible one is to some degree ultimately responsible for one's action. Ultimate sourcehood does not entail persons be 'contra-causally.' A compatibilist could consistently say an agent is a source of her behavior but not require she be an ultimate one.

Roderick Chisholm captures the incompatibilist belief of ultimate sourcehood in the following passage: "Each of us when we act, is a prime mover unmoved. In doing what we do, we cause certain events to happen, and nothing - and no one - causes us to cause those events to happen." Of course, this is typical of the noncausal variety of libertarianism I outlined in chapter two, but it is not necessarily true of incompatibilism more generally. The incompatibilist perspective understands the explanation of free action in terms of the causal chain that produces our actions to be traceable back to an agent's will and what they freely willed to do.

Kane points out that to most people this is a "literally incredible notion", but it embodies,

a central - perhaps the central - incompatibilist intuition. Of course, free agents are not prime movers in the sense of the First Cause of the cosmos. We are finite movers within the cosmos

and cannot escape being influenced by the conditions of our existence - by heredity, environment and conditioning. The incompatibilist's question is whether their[sic] is anything left over by these conditions of our existence such that our free actions can be said to be (i *) our products (i.e. caused by us) and such that (ii*) their occurring rather than not occurring here and now, or vice versa, has as its ultimate or final explanation the fact that they are caused by us.²⁰³

Chisholm's comments represent agent-causal libertarianism which Kane categorizes under the umbrella of nonoccurrent causation incompatibilism; this is in juxtaposition to his own account that he labels as a teleological intelligibility incompatibilist view. The former holds that agents have unique causal powers while the latter is content with an account of causation which does not posit agents with special causal powers. However, both camps do believe in an ideal of sourcehood. The incompatibilist understanding of sourcehood is that an agent is a producer of their free actions and that whether their action occurs or does not occur has as its ultimate explanation that the agent caused it to occur or to not occur. "Ultimate explanation" entails that an explanation of an event will have an agent be a vital reason why an event occurs or not occurs. The choice will issue from an agent's will. The agent's decision to shape her will through her choice of goals, values and ends will cause her to perform her actions. This will not imply that an agent is the sole reason for an event occurring, but that among the complex web of causal inputs the agent's decision to perform an action is a significant factor that contributes to the outcome. The incompatibilist will argue that this makes agents the ultimate source of their actions because the action performed by an agent will be attributable to them.

Consider running a small business with a partner: you and your business partner are both ultimate decision makers but individually are not the sole ones. Perhaps you and your partner do not consult each other on every decision and every action of the business. Suppose you run a

²⁰³ Robert Kane, "Two Kinds of Incompatibilism," *Philosophy and Phenomenological Research* 50, no. 2 (1989): 219-54.

restaurant and your partner takes care of the hiring while you focus on the accounting. In a sense, each of you make a difference to what the business does in that aspect of the enterprise, but the business as a whole is run by both of you. Presumably, major decisions like buying a second location will involve both you and your partner. In that case, it is clear that the business as a whole will depend on both of you such that neither of you is the sole source of that decision but individually each makes a difference to the outcome.

By analogy, human agency works within the confines of nature. Human agents are ultimate but not sole sources of their behavior as they have to work with what nature has endowed them in terms of their capacities, powers, and abilities. Insofar as these are given to human agents by nature, human agency works alongside the confines of nature to shape itself. Nature is like a partner in that it makes a difference to the abilities we have while agents make a difference to what occurs with those abilities.

To allude to themes to be discussed later on, imagine co-authors working on a novel: certain characteristics of characters are developed by one author while others are developed the other. They are each ultimately responsible for those personal characteristics they develop. They are ultimately responsible for what they produce together as one author depends on the other for details to develop the character and novel in their own way. Suppose Sara and Maria are such co-authors: Sara decides that the protagonist will have blonde hair, green eyes and be tall while Maria decides what circumstances they will be born into. Based on upon Maria's decision for the protagonist's upbringing, Sara will decide how she will respond and move forward with her life. They are together responsible for what the protagonist becomes and the narrative they are constructing.

It is helpful to meditate on ultimate responsibility as it is the fundamental idea behind ultimate sourcehood. I am going to work with Kane's definition as he lays out a formal systematic view that is quite profitable to work with:

(UR) An agent is personally responsible for E's occurring in a sense which entails that (R) 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 a arche (or sufficient ground or cause or explanation) for X, then the agent must also be personally responsible for Y.

To be sure, this is a demanding notion, but nothing less will satisfy incompatibilists. There is lots to be said about and in defense of this idea. I will from here on refer to ultimate responsibility as UR. There are clearly two parts to the condition: Responsibility (R) and Ultimacy (U). R is defined in such a way that it is open to either a compatibilist or incompatibilist interpretation. For instance, the agent voluntarily doing otherwise could be interpreted in a hypothetical sense: if the agent had wanted to act otherwise, they could have. Contributing to E occurring can also be compatibilist friendly as an agent's reasons and acting for those reasons satisfies R. 'R' is a causal notion meaning that agents are causally responsible because they produce E in the way specified in the above passage. The condition also speaks about "personal" responsibility, and the proper way to understand personal responsibility is in terms of causal responsibility. In terms of content, E can be one's character/motives/whatever. So, we can say that being personally responsible for E means you are causally responsible for your character and motives. To be personally responsible for your character and motives, there

must have been something you could have voluntarily done (or omitted) at some time or another that made or will make a difference in what you are (or in the character and motives you now and will have). This idea can also be spelled out in compatibilist terms as one could say you satisfied R by exercising whatever form of compatibilist agency one prefers.

UR, in Kane's view, is relevant to debates around alternative possibilities, and the compatibility question (from chapter two). The compatibility question straightforwardly asks if free will is compatible with determinism. The issue raised by this question is whether we need to have actual alternative possibilities (AP) when we make a choice for it to be a free choice. We can state it as: the AP requirement for free will as well as moral responsibility is that an agent must be able to or have been able to do otherwise to have free will and be morally responsible. UR is relevant here because of the R clause which is about being responsible and this includes moral responsibility. The AP requirement is open to being interpreted along compatibilist lines as part being responsible in the sense within R and AP is voluntarily doing otherwise, which is compatible with determinism. For instance, a compatibilist could say of R that if one had voluntarily wanted to do otherwise, then one would have done so. R states a minimal condition of voluntary contribution by agents toward making themselves what they are. To do this according to R, one must not be coerced or compelled but coercion and compulsion are not spelled out in compatibilist or incompatibilist terms. R, one could say, is necessary but sufficient because we could imagine scenarios where agents perform an undetermined action and so technically could have done otherwise, but they will not satisfy U. This is because, while we might be able to do otherwise in the sense of not being determined, if our character, ends and purposes were not formed by the agent then the agent won't be the ultimate source despite being able to do otherwise. Kane points out an interesting consequence of this clause, one that first

struck me as odd because this is supposed to be an incompatibilist view, which is that not every action requires an agent could have done otherwise. Upon reflection, this seems the right way to go because there are so many different causal influences and contributors to one being the kind of person one is.

4. Sufficient and Satisficing Reason

U is a sufficient condition but meeting it requires being responsible for what Kane calls a “sufficient reason,” which come in the following three forms²⁰⁴:

(1) (logically) sufficient *conditions*

(2) Sufficient causes (for example antecedent circumstances plus laws of nature) or,

(3) Sufficient *motives*. (*italics author's*)²⁰⁵

We can understand each of these as follows: Kane understands the distinction between 1 and 2 through reference to ancient doctrines of fatalism which if true could result in our actions having sufficient conditions without sufficient causes.²⁰⁶ Fatalism can be summarized in the popular saying “whatever will be, will be”, and the way to understand this in the present context as the notion of a decree of fate as a sufficient condition. Consider the ancient Greek play *Oedipus Rex*. In the play, before Oedipus is born the decrees of fate say that Oedipus will kill his father and marry his mother. The decrees of fates operate as sufficient conditions for Oedipus and the actions he will go on to perform. There are no sufficient causes for Oedipus’ actions as the laws of nature and the past are not sufficient to make this true when he is born. An example

²⁰⁴ Kane, *The Significance of Free Will*, 73.

²⁰⁵ Kane, *The Significance of Free Will*, 73.

²⁰⁶ Kane *The Significance of Free Will* 7-8.

of a sufficient cause is an action that flows from our characters and motives in conjunction with antecedent conditions plus the laws of nature. We could have a sufficient cause without motives: imagine a boulder rolling down a hill and it lands and crushes Smith and kills him. The cause of Smith's death is the boulder is sufficient as it ends his life quite decisively, but it is not necessary as Smith could still have perished in another way.

Sufficient motives are beliefs and desires which move you to act in a given set of circumstances. The difference between sufficient causes and sufficient motives can be understood by supposing that even if an action is undetermined and therefore lacks a sufficient cause, your motives were such that there was only one reasonable action you were motivated to perform. In his classic paper, "Ifs and Cans" J.L Austin imagines himself trying to hole a three-foot putt to win a golf game.²⁰⁷ Suppose that Austin's holing the putt is undetermined and so whether he succeeds or fails is uncertain. However, his character and motives plus antecedent conditions and the laws of nature are sufficient to make holing the putt the only reasonable action. If Austin misses the putt, it will only be an accident and not something he could intentionally do. This illustrates the distinction between (2) and (3) because even if an action is undetermined and lacks sufficient causes there could still be sufficient motives.²⁰⁸

So, when we discuss sufficient conditions, we can focus on a broader sufficient reason. Following the discussion on sufficient reason, we are led to some important insights about the kind of free will and moral responsibility associated with UR. The first is that indeterminism alone is not enough to deliver ultimacy. Recall in chapter two that indeterminism was

²⁰⁷ Austin, "Ifs and Can's."

²⁰⁸ There is a long-standing issue in the history of philosophy of action about the status of motives as causes. I am taking no sides in this debate and remain neutral.

conceptualized in terms of a process that involved agents overcoming an obstacle introduced by indeterminism that must be overcome if they are to achieve their goals. Indeterminism per se is not an obstacle but rather the alternative possibilities introduced by indeterminism. If an agent engaged in this process has only one goal they aim to achieve then doing otherwise will not be an achievement, it will be merely accidental. The Austin case in the last paragraph operates along these lines. Austin has one intention, to hole his putt, and has no wish to do otherwise. If Austin does otherwise, it will be merely accidental or unintentional as his intention is to be successful and not to intentionally miss his shot.

Cases like Austin's describe situations where Kane says an agent's will is 'set'. In cases where an agent's will is set, their mind are focused on only accomplishing one end, goal or purpose, and the reasons that motivate them are conclusive. To do otherwise in these situations would be accidental, like pushing the button on the pop machine for regular cola when I meant to get diet. Cases like these demonstrate that it is not enough that we can do otherwise to say that we act freely. There is little value in such a conception of freedom if all of one's alternative possibilities are mere accidents. What is needed, it seems to me, is that an agent be able to do otherwise on purpose, voluntarily and it be reasonable to do so. This is why indeterminism per se is not a sufficient condition by the standards of my libertarian view. It is a necessary condition to be able to do otherwise, but if all that indeterminism did was make it possible that I might perform an action accidentally, this would fall well short of freedom. When an agent performs an undetermined action for them to be able to freely perform alternatives they must be willingly, intentionally and reasonably able to act whatever alternative they choose. We are able to act and refrain from acting either way we act, and whatever we decide to do will be done in accordance with our wills, for reasons, and will be what we tried or wanted to do. In essence, undetermined

actions with these robust alternative possibilities make one as much an ultimate source as one can be since one has all the features that both compatibilists and incompatibilists would ascribe to free will. In terms of the indeterministic framework I have described, an agent is trying to bring about conflicting goals that serve as an obstacle to one another.

In the framework I am developing, agents are able to do otherwise not only literally but meaningfully because their reasons for acting are not conclusive, but instead are what Kane calls ‘satisficing’. A reason is satisficing when it makes an alternative possibility good enough to pursue without making the alternatives unreasonable or not also equally supported by good enough reasons. Imagine Julia really wants to go to law school and she really wants to go to graduate school. She is deeply interested in both without even thinking about it, and if she were to reflect on it this would still be the case. Julia is ‘of two minds’, and it is situations like hers that illustrate satisficing reasons because her psychology inclines her towards both options, but they support making choices which conflict with one another. Julia’s character, mind, motives and reasons do not lead her to be more attracted to one alternative over another. Kane talks about this distinction between satisficing reasons and sufficient in terms of decisiveness or conclusiveness.²⁰⁹ If I go to a coffee machine and I want a cappuccino and nothing else, my reasons are sufficient and conclusive; I have my mind made up. If I accidentally hit the button for a flat white, it would not be something I did for the reasons I had. A lot of the time we have reasons which are sufficient or conclusive in our everyday life. One’s preferences often support one option over another both non-reflectively and reflectively. The decisiveness of a reason does not require that we have to reflect on it to see its decisiveness as one could simply have no interest in any alternative. The lack of interest in an alternative counts as decisive just as much as

²⁰⁹ Robert Kane, “The Problem of Free Will: A Libertarian Perspective.”

if we were to think about the options and judge one option to be superior to others. When our minds are not so decisive but torn because we actually want to pick multiple options which conflict with each other, we have reasons that are satisficing.

I hypothesize that the best understanding of this model of deliberating about satisficing reasons is indeterministic. I will argue for the claim that this deliberative model about satisficing reasons is best understood as indeterministic in the next section. On this view, agents have genuine alternative possibilities but indeterminism should not be mistaken for the reason an agent does what she actually does. Indeterminism makes it possible for an agent to do otherwise, but it is their reasons and motivations that produce their actions. If Julia decided to go to law school, her reasons were good enough to go but so were the reasons to go to graduate school had she chosen that alternative. Hence, when she chooses, she will choose for reasons that are meaningful and motivating to her without either option being decisive or stronger than the other. Whatever she does, she will be acting for satisficing reasons and because of those reasons. Indeterminism broadens her future to more than one possibility but which future she selects will be done intelligently, intentionally and on purpose. All of the genuine possible futures open to Julia come from features of her like her reasons, character, and motives. Julia's identity is not sufficient to rule out the alternative and in fact it makes the alternative possibility attractive. Indeterminism only makes Julia's past not limit her to only one possible future; it is still Julia's reasons which end up deciding which possible future becomes the actual one. One important implication of satisficing reasons is that when they are present, we endorse the outcome of our deliberation no matter what we choose because it is what we wanted to do, we do it willingly and we do it on purpose. When we add indeterminism to satisficing reasons, we are able to genuinely do otherwise and not just by accident or unwillingly because the choice is undetermined and

done for reasons. Satisficing reasons are important because if we are to be the ultimate source of who we are, then the choices where we shape our identities have to involve reasons that are satisficing and not sufficient. This is because if we always had sufficient reasons to shape our identities, we could never choose to shape them in an alternative way that was intentional and meaningful. To be an ultimate source requires an agent be able to choose how to shape their will in different possible ways purposively, voluntarily, and reasonably.

I do not think we never have sufficient reasons or always need to possess satisficing reasons. Since every action has a sufficient reason, the ability to perform free and responsible actions in the sense of UR requires, we are responsible for shaping the sufficient reason in a specific sense. As I have said before, we need not be solely responsible for all sufficient reasons. Suppose one sufficient reason for our actions is our characters. What would be required by the standards of UR is that is we are responsible for our characters as they are a sufficient reason. Being responsible for our characters would entail that our decisions that shaped our characters made the difference to how our characters turned out. There must be actions performed by us which are produced because of our characters but our characters did not generate reasons for action that were sufficient or conclusive to act. If one's character is sufficient to produce the reason for which we act, this would mean we could not do otherwise other than by accident or contrary to the reasons for action. UR requires that we be able to shape our characters otherwise and do so intentionally, on purpose and reasonably. This entails if we shape ourselves one way, we could have shaped ourselves otherwise along the lines I just mentioned. SFAs are choices which have effects upon our characters and how they turned out. In order to be responsible for our characters and make choices like SFAs, we have to make choices for satisficing reasons and not sufficient ones. Agents must shape their characters for satisficing reasons because sufficient reasons only

allow them to shape their characters one way. If we only ever acted for sufficient reasons, there is an important sense where we would not be able to otherwise separate from the truth of determinism.

Not every action needs to have satisficing reasons instead of sufficient ones, only those actions where we directly shape our wills. These actions are Kane's SFAs, which he also calls 'will-setting' actions because we shape or incline our wills one way. When we make an SFA, the alternative possibilities are open to us not only because of indeterminism but also because every SFA has satisficing reasons. This is an important fact because it points to something important about SFAs, which is that, "*every undetermined self-forming choice is the initiation of a novel pathway into the future, whose justification lies in that future and is not fully explained by the past*. In making such a choice, we say, in effect, 'I am opting for this pathway. It is not *required* by my past reasons but is consistent with my past and is one branching pathway my life can now meaningfully take.'"²¹⁰ (author's italics) To be ultimate sources of who we are and who we become, we must make such will-setting self-forming choices. These choices will have satisficing reasons which are good enough for agent to act upon for the particular choice supported by the particular satisficing reason.²¹¹

5. The Development of Ultimacy

The account of reasons in the last section held that when we make SFAs our reasons for acting are not sufficient but satisficing. What this means is no alternative possibility is strong enough to

²¹⁰ Robert Kane, "The Complex Tapestry of Free Will: Striving Will, Indeterminism and Volitional Streams," *Synthese* 196, no. 1 (2019): 25.

²¹¹ The arguments I have offered here also serve to respond to the concern that every SFA could actually be determined. The first SFA is quite limited in our responsibility for it but the reason that every SFA cannot be determined is that this would undermine the picture of an open future where we choose what to do for satisficing reasons. The first SFA might be heavily shaped by factors beyond our control but it is mistaken to believe they could all be so shaped. For if we are self-forming agents, our pasts cannot limit our future to one single pathway.

make the other unreasonable or unattractive to the agent. This opens the way to an understanding of self-shaping involving ultimate responsibility and SFAs as a process of agential development. An agent's character and motives in SFAs are not sufficient to rule out alternatives so they only engender satisficing reasons. When an agent acts for a satisficing reason, they develop their wills in such a way as to make them more sufficient and thereby make reasons become sufficient. I am going to discuss this process of development as it leads to understanding how agents are ultimately responsible for who they are and what it means to say one is an ultimate source of one's personal identity. The account will emphasize the gradual nature of the development of ultimate sourcehood as well as the fact that ultimate sourcehood involves both compatibilist and incompatibilist senses of freedom, control and responsibility. As ultimate sourcehood does not require sole sourcehood, this should not come as a surprise. Incompatibilism on the model offered by my thesis sees incompatibilism growing out of compatibilism.

Consider the following case from John Lemos of an agent called Rosa:

Imagine Rosa who is quite young – three or four years old. She has been told not to eat from the cookie jar unless she has her parents' permission. She has already had her daily allotment of cookies, but now she wants more. She sees that she might be able to sneak a cookie from the jar while no one is looking, but she fears she may be caught and reprimanded. She's torn, and she wonders what she should do. Suppose this is the occasion of her first SFA. If so, then she tries/wills to (A) sneak a cookie and to (B) obey her parents. Let's suppose she ultimately caves to temptation and sneaks a cookie from the jar. If so, and assuming this is an SFA, then the choice to do so is causally undetermined. On Kane's view even if each of the dual willings/efforts meet compatibilist criteria of freedom, she has very little responsibility for her choice, because she has not formed the character from which her dual efforts arise; rather, the desires and reasons which motivate these efforts are at this point in her life too much a product of genetic and environmental factors. She has not yet performed enough of the free undetermined SFAs to give her much responsibility for her choice.²¹²

²¹² John Lemos, "Self-Forming Acts and the Grounds of Responsibility," *Philosophia (Ramat Gan)* 43, no. 1 (2015) (2011): 142.

Let's dig deeper into the case to see what can be learned about responsibility and control. I propose to describe Rosa pretheoretically and from a compatibilist and incompatibilist perspective. I will then turn to showing that the incompatibilist perspective captures more of our pretheoretical intuitions about Rosa. What the incompatibilist will be able to deliver that the compatibilist will not is ultimacy and I will show that this is important to the Rosa case.

To start with the pretheoretical interpretation, Rosa's free will and responsibility is limited because she is a child. Children obviously have limited levels of freedom and responsibility for who they are and what they do. No one would reasonably say that newborn infants have any control, freedom or responsibility for what they do. Clearly freedom and responsibility are things that develop gradually with human beings as we start out being driven by impulse, and as we mature gain more control. As children develop, their capacity to form intentions and reasons as well as respond to them makes them more appropriately thought to have some degree of freedom and limited responsibility. We also think of children like Rosa developing more responsibility not only in terms of the stage of development but also in terms of the patterns they engage in. The first time Rosa steals a cookie she has no prior experience and so it is quite natural to say she does not know any better. She has to learn not to steal cookies or steal in general. But if she continues to engage in this behavior as she develops, it becomes more appropriate pretheoretically to assign more substantive responsibility and moral blame to her. She has been told and, in a sense, given the opportunity not to do it again and she persists despite this. This fits with the earlier point that children need to be educated regarding what to do and what not to do. As the case is presented, Rosa is able to recognize the reasons to not steal the cookie, but she also has reasons to do it in spite of the other compelling reasons. She has a level of development where children can respond to reasons which pull an agent in opposite

directions. Now, if Rosa has done this before and so she knows it's wrong to steal, Rosa is beginning to emerge as a clear agent who is an independent source of activity. She is developing to become a full agent as she shapes herself through the actions and choices she makes. Rosa is developing into an agent and a person through the development of her history. What I mean to say is that her moral responsibility for her actions is quite limited the first time she performs an action. But the second, third, and fourth instances of the same action start to create a pattern and history that incline her. The maturation of children and their development is comprised of learning something is wrong and to not do it again. And when they do repeat whatever action it might be, they are held more responsible because they were already made aware of what they should have done; they also are more capable of listening and learning through exercising their capacities.

As Rosa grows up and develops into a full-blown agent the relationship between her and her movements changes into one where she is making decisions and she is the source of what she does. The first time Rosa steals a cookie it is intuitive to say she is a source of her behavior, but it is also intuitive to say, her sourcehood and responsibility is quite limited. As Lemos points out, Rosa's efforts and motivations are still very much the product of her environment and her genetics and not a character she has shaped. The relationship between Rosa and her cookie stealing is crude due to her immaturity. As Sourcehood is a matter of degree, Rosa is a source in a very limited way. Her capacity to respond to reasons is too crude to be a ultimate source but it is sufficient to a source. The important point here is that immature children often just act instinctively and respond to whatever reason is the strongest. Rosa's situation is unique as for the first time she is able to be aware of different reasons competing with each other. Her sensitivity to reasons is at a point where she can begin to respond to different compelling motivations.

The nature of the relationship between Rosa and her reasons is one where they do not express anything about Rosa. The source of the reasons is not Rosa's character because she has not formed it. She instead is motivated by her environment and genetics, and after she acts she will have begun forming her character. Through acting, Rosa changes her relationship to her reasons because by choosing to act for reasons, which might initially come from external factors, she makes them hers. Her awareness of herself in relationship to the reasons for which she acts changes when she makes a choice to act for a reason. Her limited power to choose changes the relationship to her reasons as she has now shaped her will to a degree. The next time Rosa is deliberating over stealing more than her allotment of cookies, she has a reason that is not wholly external to her. The first time Rosa steals the cookies, she had no prior understanding of the situation or the reasons that she is responding to. She does not see herself as an agent that acts for reasons. Rosa deliberates and just acts without a sense of self being expressed through the reasons to which she responds.

We can describe Rosa from both a compatibilist and incompatibilist perspective. Lemos writes that there is a sense in which Rosa's efforts are hers because she meets good compatibilist criteria. Lemos describes Rosa's free will and responsibility as limited but only in certain ways. She is not responsible for the character from which her motivating reasons and desires flow, but as an agent, Rosa still plays an important causal role. As Lemos writes, "It is fundamentally important to note that if she did not continue to exert the effort to sneak a cookie while simultaneously trying to obey her parents then the sneaking of the cookie would never occur. Rather, without that effort she makes to sneak the cookie, she would not do so and would, instead, opt to obey her parents. In this way, Rosa, through her own effort plays a significant

causal role in making the choice to sneak a cookie.”²¹³ Both libertarians and compatibilists should agree on Rosa playing a causal role.

Compatibilists and incompatibilists can and should both agree young children like Rosa have very little if any responsibility or free will. A theorist attracted to characterizing Rosa in terms of compatibilism will interpret the case along the lines of their preferred view. A compatibilist like Fischer or Frankfurt could say that Rosa lacks responsibility or freedom because she lacks the necessary sensitivity to reasons or the capacity to reflect and endorse her will. Suppose Rosa had just enough of a capacity to reflect on and respond to the inputs from her environment. It is not the same degree as an older child or an adult but enough to judge her as making a choice. Sensitivity to reasons is open to being a matter of degree, and we could say Rosa is sensitive to reasons minimally such that she makes a choice but has no or minimal responsibility for what she does. It is enough to perform a free action but not enough to appropriately assign moral desert. I think what compatibilists like Fischer and Frankfurt could say is that, as Rosa matures, her capacity to be sensitive to reasons would develop and eventually it would be appropriate to assign moral blame or praise to her. Watson’s particular theory would say that the lack of harmony between her valuational system and motivation system would be the missing component, and this is what is to be expected of a child. Rosa has the capacity to form a valuational system, but the capacity will not be exercised until she reaches an age of maturity to do so.

Compatibilists sensitive to history like Wolf, Mele and in some respects Doris might say that Rosa’s history does not yet lend itself to her being morally responsible. Mele’s account

²¹³ Lemos, “Self-Forming Acts and the Grounds of Responsibility,” 142.

would say Rosa's responsibility depends on whether her history undermines it. Recall that it is not determinism per se but certain deterministic histories which threaten free will and moral responsibility. If Rosa is raised in an unhealthy environment and learns bad habits from bad role models, then this would have the potential to stop her from being responsible. This is along the same lines as Wolf's JoJo case. JoJo learned from his ruthless father who taught JoJo that the right way to treat others is with cruelty and he was so inculcated by his upbringing that it prevents him from being able to respond to moral reasons. He has a deterministic history that undermines his freedom and moral responsibility. Carolina Sartorio also says something along the same lines as Wolf. If a person was abused as a child and this event determines this person to make a bad choice, this would be relevant to judging our blameworthiness.²¹⁴ She takes the same line as Mele in claiming that being determined to act by a specific event in one's life is quite different from being determined in a general sense by nature.

Doris's view of agency holds that exercising agency and attributing responsibility is applicable to both discrete actions and temporally extended processes. Rosa's choice to steal a cookie is a discrete unit but if her activity were to be of a kind that took place over time, like a career, then this would be an appropriate description of her too. Doris' discussion of Huck Finn emphasized that Huck's responsibility was in terms of the expression of his values, which are for Doris linked to the exercise of agency over time. Due to Rosa's immaturity, she lacks values. This is related though separate from history because Doris' view of agency believes we are able to exercise our agency at a given time and over time. Agency is exercised diachronically and

²¹⁴ Carolina Sartorio, "Free Will and Determinism: A Compatibilism," In *Do We Have Free Will?*, 1st ed., (Routledge: 2022).

expresses our values. Rosa might act diachronically but she lacks values so she her agency is not fully exercised.

An incompatibilist reading of Rosa can accept much of the compatibilist interpretation. Given what I have said about compatibilism, any incompatibilist theorist attracted to my sort of libertarianism must accept that compatibilism accurately describes certain features of agency. A morally responsible agent must be sane, reasonable, capable of acting in accordance with their will and expressing their values. An incompatibilist interpretation of the case will make Rosa's decision an undetermined one. Lemos already offered one reading of the case, but we will avoid the use of SFAs to avoid the risk of begging the question. Since we are postulating, for the sake of argument, that Rosa's choice is undetermined, she has actual alternative possibilities. Recall from chapter three that compatibilism could attribute alternative possibilities in a hypothetical sense: if one had thought something different or the laws of nature were different then you would have done otherwise, all else being equal. On the incompatibilist's construal of Rosa, she has actual or categorical possibilities which means in the actual world, and not a separate possible world, she can genuinely do otherwise. Rosa has a genuinely open future when she makes her decision. The openness of the future means it is up to Rosa which fork in the garden of forking pathways she chooses.

A salient feature of the incompatibilist interpretation of Rosa is ultimacy. Rosa develops a relationship to her actions where she becomes an ultimate source. Intuitively, children are excused for the actions that they perform because it is not appropriate to attribute the actions they perform in a required sense. They literally do things, but they act from impulse or ignorance and in other ways that make it inappropriate to attribute what they did to them in such a way to attach moral blame to them. Rosa's case describes a state of development where she acts for reasons

and is just beginning to come into her own. Lemos' characterization of Rosa in the case brings this out "[Rosa] has very little responsibility for her choice, because she has not formed the character from which her dual efforts arise; rather, the desires and reasons which motivate these efforts are at this point in her life too much a product of genetic and environmental factors." She might be acting for reasons, but her capacity to form and be sensitive to reasons is too crude because it is undeveloped. To say Rosa is too much of a product of genetic and environmental factors is to say her reasons are not really her own because she does not stand in the right relationship to them. In Rosa's context, the reasons which motivate her to act come from the external factors that Lemos identifies. Only as Rosa continues to exercise her ability to be sensitive and respond to reasons will those reasons become attributable to her, and hence, her actions will be attributable to her as an ultimate source of them. This tracks with the development of Rosa gradually becoming ultimately responsible as she develops into an agent that is ultimately responsible and an ultimate source.

I think the way to understand the issue of Rosa's relationship to her reasons is through understanding that the situation she is in is one where she is "getting in her own way." Her identity and reasons pose obstacles for Rosa. Indeterminism is a necessary condition for Rosa's identity and reasons operating in this manner. Rosa's pursuit of A is interfered with by attraction to B, and she is genuinely able to choose either. As Rosa does not have a formed character, the reasons that are obstacles in her case come from the external factors like her genetics. Through the performance of the action to steal or not steal the cookies, Rosa's choice shapes her such that she now has a degree of a self that generates reasons for action. Rosa's reasons are not ones she has authored, and this is the key difference in the relationship between her before and after her first undetermined choice. She has efforts that are hers in the terms of

being a causal contributor, but then after she makes a choice, she becomes a source as the author of her reasons and efforts.

Given that my account adopts certain compatibilist features, readers might wonder why we should favor incompatibilism in describing Rosa. Compatibilists do not see ultimacy as important while incompatibilists hold that ultimacy is fundamental to free will and moral responsibility. Ultimacy, as defined in UR, says that we must be responsible for the sufficient reason, explanation or cause for acting. It is clear from the discussion on satisficing reasons play an important role in ultimacy. In situations where we have satisficing reasons, we are able to do otherwise, and when a decision which is supported by satisficing reasons is undetermined I argued that what occurs or not occurs is ultimately up-to-us. While indeterminism makes it possible for agents to do otherwise, it is the satisficing reasons that explain why they do otherwise. This is worth emphasizing because compatibilists will disagree that indeterminism is necessary, but it is not doing the bulk of the work in the present account. However, indeterminism is only an ingredient not the whole recipe for freedom. The role it plays is in making an agent face a metaphysically open future. The next step is where satisficing reasons takes over and satisficing reasons don't assume determinism or indeterminism. The conjunction of indeterminism and satisficing reasons together deliver ultimacy. Given a metaphysically open future with divergent possible worlds, an agent is able to do otherwise because her reasons support the different alternatives.

Acting for reasons is necessary for ultimacy. This connects incompatibilism to compatibilism because compatibilist accounts like Fischer's give significant weight to acting for reasons. If we lacked the ability to properly respond to reasons, then we would fail to meet the requirement of ultimacy. A compatibilist could ask: "If certain compatibilist criterion are correct,

why should we accept an incompatibilist interpretation of Rosa?” What the incompatibilist describes about Rosa which the compatibilist leaves out is a certain characterization of the relationship between Rosa and her life as she become an agent. What compatibilism leaves out is Rosa becoming the author of her life. When Rosa makes her first choice, she changes the manner in which she stands towards herself and her life. Through the first choice of the sort Rosa is making, Rosa is authoring her will and her life. Through this novel authorship, Rosa acquires a unique sort of ownership of herself. Prior to this sort of choice, Rosa’s character comes entirely from her environment and genetics. When she acted, she was not the author of the efforts or character. In the sense that she played a causal role in the production of the event, it is hers. However, through making an undetermined choice. Rosa is not only a causal contributor, but she is also to a degree the source or author of her efforts and character.

I think compatibilists are right to say that free will and moral responsibility are related to reasons and being sensitive to them, but what they leave out that incompatibilism captures is this sensitivity developing in such a way that opens the possibility for agents to responding to different compelling reasons at the same time. Given how Rosa’s identity is constituted at the time, she is sensitive to certain considerations that generate reasons that compel her in incompatible ways. Lemos says Rosa’s character is too much a product of her environment and genetics. In her first undetermined choice with rationally satisficing reasons, these factors have their sourcehood lessened while Rosa’s choices become greater sources of who she is and what she does. Her identity is becoming less the product of her environment and more the result of Rosa’s authoring it through the satisficing reasons she chooses for; with indeterminism Rosa is able to genuinely choose between her satisficing reasons. She wants to do A but she is attracted

to B and this attraction gets in the way, and indeterminism makes it possible for her to choose either A or B.

The incompatibilist model drawn in this way shows Rosa is not limited to being able to genuinely choose only one option. Each option has reasons Rosa is sensitive and responds to, but compatibilism leaves out the desirability of Rosa having an ability to genuinely choose between options that rationally satisfy. Compatibilism can make use of satisficing reasons, but the doctrine does not see the relevance of doing otherwise. The relevance, I think, comes from the perspective of Rosa taking a perspective of authorship of her life. In the first moment she makes the right sort of undetermined choice, Rosa's character is not determined to one possibility by her genetics and environment. It is open to Rosa to choose given the inputs and the satisficing reasons whether to choose A or B. It is a better interpretation of Rosa to say she can do otherwise because her power to choose alternative possibilities shows the character she is starting to shape and form is really grounded in her and her choice and not external necessitating factors. She takes and acquires partial ownership, and her ownership will increase through the maturation process as she makes more of these undetermined choices for satisficing reasons.

When Rosa makes an ultimately free choice, she does meet certain compatibilist criteria of acting for reasons, but she also does more besides since her sensitivity to reasons opens the possibility of acting for different reasons at the same time. Her sensitivity to reasons develops from the exercise of the capacity because she begins to get better at responding to reasons. The development of the sensitivity to reasons opening the possibility of doing otherwise in certain situations with competing and satisficing reasons captures the development of agency and this process of development would seem to deliver ultimacy. Rosa's sensitivity to reasons makes her motives support acting in different contrary ways in the same context and which decision is

made is ultimate up-to-her; in other words, the development of her sensitivity to reasons resulting in her recognition of different reasons making it possible to do otherwise, assuming her choice is undermined, will make her the ultimate source of her choice.

Whatever she does, Rosa is the ultimate source of what occurs or does not occur because she is able to choose between different forking pathways. She is not limited to only one possible pathway. This is because of both indeterminism and her satisficing reasons. If Rosa did not have satisficing reasons for what she does, then even though her choice was undetermined she still could not have done otherwise in a sense where she is responsible. She would only do otherwise by accident and would not endorse the outcome as what she wanted to do. She would not be responsible for what she did because of the lack of endorsement of what occurred. Of course, there are cases where people are responsible for accidents, like an assassin accidentally killing the wrong person. But situations like Rosa's are ones where, without the robust alternatives grounded in satisficing reasons, we lack the ultimacy needed to be responsible in the sense of UR.

UR is something that accumulates over time. Rosa is just beginning to be able to be sensitive enough to the reasons she is responding to. It is a skill that needs to be exercised for Rosa to get better at for her to become more sensitive to reasons in the right way. Consider what might happen if Rosa were to disobey her parents and steal the cookie. The next time her parents tell her no more cookies, she might be more inclined to do it again. The more Rosa steals cookies, the more she shapes her will and cultivates a pattern of behavior that generates a sufficient reason from her point of view to do what she does. She will become ultimately responsible for her sufficient reason because it emerges from a process where she acts for satisficing reasons and we can trace the origin of the sufficient reason to Rosa as the ultimate source of it. Rosa makes

the difference as to what sufficient reason she has since the choices she makes that produce a sufficient reason involved alternative possibilities. She is the ultimate source of her will and ultimately responsible for it.

Ultimacy is also a matter of degree. Rosa's parents play an obvious role in the formation of her will. They also play a role in her being an ultimate source and ultimately responsible for what she does. How Rosa's parents respond to her behavior provides feedback that informs her future deliberations, reasons and sensitivity to those reasons. If Rosa's parents never discipline her when she acts contrary to their commands, Rosa will have little reason in the future to believe she should listen to them other than when it suits her interests. She will end up shaping her will such that when her desires conflict with the requests of other agents, she will not be inclined to respond to those reasons. If Rosa's parents were to implement discipline grossly disproportionate to her action, she will be in the same boat as JoJo whose father taught him to respond to the slightest infraction with barbarous cruelty. It is quite clear that Rosa's decisions play a significant role in shaping her will, but the reasons for which she acts are heavily influenced by the interaction she has with other people in important ways.

Rosa demonstrates how ultimacy in a general sense emerges through the engagement of one's environment. The case shows how the baseline capacity to respond to reasons that compatibilists stress leads to an incompatibilist kind of free will and moral responsibility because, as persons become more able to respond to reasons, their sensitivity is greater and allows them to respond to different kinds of reasons at the same time. I think our capacity to become more sensitive to reasons is more plausible in an undetermined world. As a very young child, Rosa's sensitivity to reasons is quite crude, but as she develops and matures, making choices that lead to patterns of behavior, she is developing incompatibilist kinds of control,

freedom and responsibility. This provides a suggestive case for preferring the incompatibilist interpretation of Rosa and the development of our capacities for choice over the compatibilist one. We therefore have an account of ultimacy as a gradual process that emerges slowly and respects the bounds of nature. This account seems to me to give good reason to think incompatibilism more accurately describes the development of agency in the Rosa case and in a more general sense. Of course, the incompatibilist view here is not empirically sound as it is an open scientific question whether determinism is true. What I have done here is provide a case for the conceptual plausibility of ultimate libertarian sourcehood.

6. Responsibility and Ultimacy

There are different sense of ‘R’. In chapter two, I outlined three senses of freedom associated with the phrase ‘acting of one’s own free will,’ and it would be useful to similarly disambiguate responsibility. I have talked about responsibility in terms of UR, but I have also discussed responsibility more generally too. Kane has a helpful distinction between different kinds of responsibility described in terms of dimensions which seem helpful here. There are two dimensions of responsibility:

First Dimension of Responsibility: expressing the will (character, motives, and purposes) *one has in action*, and doing so *voluntarily* (without being constrained or hindered or forced), and doing so *intentionally* (knowingly and purposefully) (authors italics).²¹⁵

Second Dimension of Responsibility: To be responsible in this second dimension it must be that at least some time in your life, when you act responsibly and hence voluntarily

²¹⁵ Kane, “The Complex Tapestry,” 115.

and intentionally in the first dimension, it was possible for you to have *voluntarily and intentionally done otherwise*, not by being forced or by accident, but in a manner that would *also* have expressed the true quality of your will and the self that you were at that time.”²¹⁶

The first dimension of responsibility is the kind of responsibility that Doris advocated for in chapter three. The way Doris tells the Huck Finn case, Huck is responsible for what he does because his action expresses his values, and while Doris does not, we could say Huck expresses his will. The ability to express yourself in action is consistent with determinism as the criterion requires you are not constrained or forced and that you act voluntarily and intentionally. These criteria are compatible with being determined to act. The second sense of responsibility is not compatible with determinism. This kind of responsibility requires that one have robust alternative possibilities and that it was possible for you to actually do otherwise for reasons which are satisficing. Now, again, that we act for satisficing reasons by itself does not show this sense of responsibility is incompatible with determinism. Rather, it is that if determinism is true, there never was and never will be a time in one’s life where it is possible for you to act voluntarily and intentionally, and hence in the second dimensional sense, done otherwise. What determinism rules out is robust alternative possibilities which are different ways one could not just express one’s will but different ways one shapes the very thing that is expressed. You may shape it one way if you make a deterministic choice but it is never possible to robustly choose to shape it otherwise. You are limited to only one possible future in regard to the way you shape your will and the kind of person you thereby become. The second sense of responsibility is linked to ultimate sourcehood because an ultimate source means being responsible for who you

²¹⁶ Kane, “The Complex Tapestry,” 115.

are. Being responsible for who you are in the second dimensional sense means that it is possible to shape one's will in more than one way. This is implied by having an open future of forking pathways and which forking pathway is actualized depending upon the agent and her decision. The account of satisficing reasons offers a way of explaining how intelligently we can form ourselves and be responsible for who we are.

The nature of satisficing reasons aids in understanding ultimacy because it pins down a feature of rationality that is by now quite clear. Decisions can have multiple reasons that support different possible options with no particular set of reasons appearing stronger compared to the alternatives to the agent. Satisficing reasons are an essential aspect of SFAs. Agents who make SFAs are divided over what they ought to do precisely because their reasons are not sufficient but satisficing. One's uncertainty over what to do is seen in one's satisficing reasons which are expressive of competing parts of one's will. I suggest that each reason is a different value of the agent. We could say that an agent's satisficing reasons comes from an internal struggle in their wills because they are ambivalent in the sense Doris described in chapter three. Recall that for Doris, we are often ambivalent about our values and this ambivalence is part of our agency. My discussion of satisficing reasons provides good reason to link satisficing reason to ambivalence and thus agency. In virtue of the connection between satisficing reasons and ultimacy, we can also see the connection between agency and ultimacy. Our valuational conflict is rooted in our wills and identities. In accordance with the second dimension of responsibility, whichever satisficing reason we act for, it is agents that decide what values to shape themselves. In other words, to the extent values constitute an agent's identity, agents will be the ultimate source of them. They will be the ultimate source of their values as they will make choices for satisficing reasons; the effects of these choices will impact their values.

The last sentence might invite an interpretation of ultimacy as the kind of all or nothing thing I have said it is not, and so, it is beneficial to demonstrate how what I have outlined works as a matter of degree. This is essential as I am committed to showing that incompatibilism and compatibilism can agree on certain aspects of free agency and free will. For my libertarian view agrees with Kane that, “incompatibilist freedom and control presuppose compatibilist freedom and control. We cannot get to incompatibilist freedom in one fell swoop in the real world. That is one leap too far. We must get there step wise, by exercising compatibilist... control over cognitive processes aimed at making choices”²¹⁷

6. Fischer and Sourcehood Incompatibilism

The position for the incompatibility of free will and determinism has been broadly stated in terms of ultimate sourcehood. We can refine the view through a discussion of Fischer’s important article, “Sourcehood: Playing The Cards That Are Dealt You” where he engages with sourcehood incompatibilism.²¹⁸ Fischer goes over two sourcehood accounts developed by Galen Strawson and Saul Smilansky. The labour of going over these views allows us to see a more sophisticated picture of ultimate sourcehood libertarianism. Fischer begins by stating it is “uncontentious” that persons want to be the “initiators” or “ultimate sources” of our behavior and that morally responsible agents must meet some conditions that capture these ideas in an appropriate manner.²¹⁹ The problem Fischer raises is with the criteria articulated by source incompatibilists such as Galen Strawson, Saul Smilansky, and Kane. Fischer’s main point of

²¹⁷ Robert Kane, “Rethinking Free Will: New Perspectives on an Ancient Problem,” In *The Oxford Handbook of Free Will*, eds Robert Kane. 2nd edition (New York: Oxford University Press, 2011): 404, n.14.

²¹⁸ John Martin Fischer, “Sourcehood: Playing The Cards That Are Dealt You,” in *Deep Control: Essays on Free Will and Value* (New York: Oxford University Press, 2011).

²¹⁹ Fischer, “Sourcehood: Playing The Cards That Are Dealt You,” 164.

contention is the view of self-creation believed to be entailed by sourcehood in source incompatibilism. Fischer makes a lot of important points relevant to ultimacy. However, I argue that Kane's view does not suffer from Fischer's objections against the ultimacy requirement.

Let's start with an argument from Galen Strawson, as Fischer critiques Strawson's Basic Argument in order to raise what he takes to be problematic with an ultimacy requirement for moral responsibility. Strawson's Basic Argument runs in the following way:

(1) You do what you do because of the way you are.

So

(2) To be truly morally responsible for what you do you must be truly responsible for the way you are at least in certain crucial mental respects.

But

(3) You cannot be truly responsible for the way you are, so you cannot be truly responsible for what you do. Why can't you be truly responsible for the way you are?

Because

(4) To be truly responsible for the way you are, you must have intentionally brought it about that you are the way you are, and this is impossible.

Why is it impossible? Well, suppose it is not. Suppose that

(5) You have somehow intentionally brought it about that you are the way you now are, and that you have brought this about in such a way that you can now be said to be truly responsible for being the way you are now.

(6) You must already have had a certain nature N in the light of which you intentionally brought it about that you are as you now are.

But then

(7) For it to be true you and you alone are truly responsible for how you now are, you must be truly responsible for having had the nature N in the light of which you intentionally brought it about that you are the way you now are.

So

(8) You must have intentionally brought it about that you had that nature N, in which case you must have existed already with a prior nature in the light of which you intentionally brought it

about that you had the nature N in the light of which you intentionally brought it about that you are the way you now are...²²⁰

This is a regress argument against the very existence of free will and moral responsibility.²²¹ As the eighth premise makes clear, Strawson places a requirement on moral responsibility that one has to have made oneself the way one is without an already existing nature in place. He approvingly quotes Nietzsche who makes the point about the inherent regress in Strawson's view of free will when he writes,

The *causa sui* is the best self-contradiction hitherto imagined, a kind of logical rape and unnaturalness: but mankind's extravagant pride has managed to get itself deeply and frightfully entangled with precisely this piece of nonsense. For the desire for 'freedom of will' in that metaphysical superlative sense... is nothing less than the desire to be precisely that *causa sui* and, with more than Munchhausen temerity, to pull oneself into existence out of the swamp of nothingness by one's own hair.²²²

Strawson's argument believes that free will and moral responsibility require individuals be precisely the *causa sui* that Nietzsche believes is an incoherent absurdity. The idea of *causa sui* is understood to be a model of self-creation *ex-nihilo* which is impossible. Strawson believes this entails that no one can exercise free will or be truly morally responsible.

Fischer denies not only this entailment but the whole picture of responsibility formulated by Strawson's argument. Fischer doesn't reject self-creation or ultimate sourcehood outright. His problem is with requiring such an 'inflated' ideal of self-creation. He provides an exegesis of Joel Feinberg's account of self-formation. I won't go over all of Feinberg's observations but he does convey a relevant sense of self-formation regarding his own theory of autonomy. Feinberg states,

²²⁰ The argument is presented verbatim from Galen Strawson, "The Impossibility of Moral Responsibility," *Philosophical Studies*, 75, (1994):12-14.

²²¹ Strawson treats moral responsibility as inseparable from free will and he uses the argument to argue for the impossibility of both moral responsibility and free will.

²²² Friedrich Nietzsche, *Beyond Good and Evil* (New York: Penguin Books, 2003): 51-52.

A common-sense account of self-creation (the term “self-determination” has a less grating and paradoxical sound) can be given, provided we avoid the mistake of thinking that there can be no self-determination unless the self that does the determining is already fully formed... the extent of the child’s role in his own shaping is... a process of continuous growth already begun at birth. From the very beginning that process is given its own distinctive slant by the influences of heredity and early environment. At a time so early that the questions of how to socialize and educate the child have not even arisen yet, the twig will be bent in a certain definite direction. From then on, the parents in promoting the child’s eventual autonomy will have to respect that initial bias.²²³

Fischer believes Feinberg’s comments in the above passage contain all the criteria essential for ascriptions of moral responsibility. Fischer asks the question, “Why... is it more plausible to jettison moral responsibility and cling to a very demanding notion of self-creation (“an inflated conception of autonomy”) than to scale down the demands of self-creation to something more reasonable?”²²⁴ Fischer believes that this conception of self-creation and autonomy is more reasonable and sufficient for moral responsibility.²²⁵

Utilizing Feinberg’s account as an example of a noninflated view that allows one to be morally responsible without being a *causa sui*, Fischer begins his critique of Strawson. Fischer begins his analysis with pointing out how pervasive luck is in terms of persons becoming agents. He says, “I had no control over whether I was born with a significant brain lesion that would impair or expunge my agency. Had I been born with such a lesion, I would never have developed into an agent at all, or would have developed into an agent with a very different character and set of dispositions.” He continues

²²³ Joel Feinberg, *Harm to Self*. Volume Three: *The Moral Limits of the Criminal Law*. (New York: Oxford University Press, 1986): 33-34.

²²⁴ Fischer, “Sourcehood: Playing The Cards That Are Dealt You,” 168.

²²⁵ I should say that since Fischer is a semi-compatibilist he believes that we have the kind of free will to be morally responsible, it’s just that this notion of free will doesn’t include the freedom to do otherwise.

When one begins to think about this sort of thing, one quickly realizes that we are incredibly lucky to be as we are. I had no control over the fact that I was not hit by a bolt of lightning when I was young... Life is extraordinarily fragile, and ... we are remarkably lucky to be agents at all... We are not “ultimately responsible” for “the way we are,” and yet it just seems crazy to suppose that we are thereby relieved of moral responsibility for our behavior.²²⁶

Fischer’s argument then turns to what he thinks is a problematic notion of control that lies at the heart of Strawson’s view. Given that there are so many factors beyond our control necessary for individuals to be agents at all, Fischer believes Strawson holds a requirement of “total control” for free will and moral responsibility. Total control entails, “the locus of control is entirely within the agent; there is no factor external to the agent (and out of the agent’s control) that is a causal contributor to the outcome in question and which is such that if it were not to occur, the outcome would not occur. In this view, it is as if we (*qua* moral agents) have a protective bubble around us, or perhaps we are in our armored vehicles.”²²⁷

Such a picture of agency and autonomy is both inflated and illusory. Strawson could very well reply this is exactly his point. Reflecting on the requirements for free will and moral responsibility leads one to the conclusion that what would be needed for their exercise is impossible. Fischer replies that the sensible conclusion is not that free will and moral responsibility are impossible, but that Strawson has inflated their requirements. A full examination of Strawson would be needed to rebut his theory but Fischer does not provide one. The minimal point that he wishes to establish is only that Strawson’s criteria of responsibility for the ‘way we are’ do not survive critical scrutiny.

²²⁶ Fischer, “Sourcehood: Playing The Cards That Are Dealt You,” 168-169.

²²⁷ Fischer, “Sourcehood: Playing The Cards That Are Dealt You,” 171.

Fischer critiques another sourcehood incompatibilist argument offered by Saul Smilansky. Smilansky's view is described as 'Illusionism' and I find it interesting but not convincing.²²⁸ Smilansky believes free will does not exist: it is an illusion but it is a fiction worth maintaining.²²⁹ Smilansky has two aspects to his view and Fischer deals with only one of them so we will only analyse in depth one part of Smilansky's view.²³⁰ Smilansky believes that one should adopt a dualistic approach to the free will problem. By 'dualism' he does not refer to the well-known mind-body theory of the same name that posits the existence of physical and non-physical substances interacting with one another. For Smilansky, 'dualism' here means adopting both hard incompatibilism and compatibilism. Smilansky labels this hybrid of these two opposing views "Fundamental Dualism."²³¹

Smilansky argues that a harmful assumption of 'monism' has prevented the development of his fundamental dualist proposal. The assumption of monism is, "the assumption that, on the compatibility question... one must affirm compatibilism or incompatibilism. In fact, there is no conceptual basis for thinking that the assumption of monism is necessary... it is possible to hold a mixed, intermediate position that is not fully consistent with either."²³² Smilansky asserts there are valid aspects to both compatibilism and hard determinism and outright rejects libertarian free

²²⁸ Saul Smilansky, *Free Will and Illusion* (Oxford: Clarendon Press, 2000).

²²⁹ Saul Smilansky, "Free Will, Fundamental Dualism, and The Centrality of Illusionism" in *The Oxford Handbook of Free Will*, 2nd ed, eds Robert Kane, (New York: Oxford University Press, 2012): 425-441.

²³⁰ The part of Smilansky I will not be engaging is his illusionist proposal. Smilansky argues that if we accept his fundamental dualism then we will indirectly be led to the thesis that some kind of dualism must be maintained. He believes that if society accepted the belief that we do not have libertarian free will and went all the way with hard determinism society would be seriously harmed. In order to maintain a community of responsibility where individuals are accountable for their actions the truth of no libertarian free will should be kept inside 'elitist' institutions. Compatibilism will only work to keep the community functioning if the general public does not see their libertarian assumptions are false and that everything they do is an unfolding of the given.

²³¹ Smilansky, "Free Will, Fundamental Dualism, and The Centrality of Illusion," 427.

²³² Smilansky, "Free Will, Fundamental Dualism, and The Centrality of Illusion," 427.

will. Compatibilism can make some sense of free actions but Smilansky charges it as morally shallow for reasons that vindicate the hard determinist. He says, “We can make sense of the notion of autonomy or self-determination on the compatibilist level but, if there is no libertarian free will, no one can be ultimately in control, ultimately responsible, for this self and its determinations. *Everything* that takes place on the compatibilist level becomes on the ultimate hard determinist level ‘what was merely *there*,’ ...”²³³

Smilansky’s criticism of compatibilism is that, from what he calls the ‘ultimate perspective’, every action which is free by compatibilist standards is the ‘unfolding of the arbitrarily given’. All actions and choices are the necessary result of what came before and individuals have no ability to choose otherwise than they do. Agents choose as they do because of the already given character and motivation set which they have as a result of pure luck. Thus, they are not responsible for who we are and what we do.

Smilansky synthesizes his complaint with compatibilism and his charge of moral shallowness when he writes,

After all, we are also all fortunate that the earth was not hit yesterday by a huge meteorite, but that seems scarcely relevant to the free will issue. Luck is undoubtedly present in our lives, but the central question is how its presence is manifested... (T)he difficulty is that compatibilist control is set by the way we are constituted so that every choice we make and every action we undertake is an unfolding of the arbitrarily given-the luck is not located in some corner but, when we look deeply, we see that it *goes all the way through*.²³⁴

Fischer’s response to Smilansky draws out something important. He claims we need not be in control of necessary conditions such as the sun continuing to shine or the meteorite not hitting the earth. These conditions are necessary, but they are also all a matter of luck. Fischer’s point

²³³ Smilansky, “Free Will, Fundamental Dualism, and The Centrality of Illusion,” 428.

²³⁴ Saul Smilansky, “Compatibilism: The Argument from Shallowness,” *Philosophical Studies* 115 (2003): 275.

then is not that luck is in some irrelevant corner but pervades our daily lives. Fischer zooms in on Smilansky's phrase '*luck goes all the way through*' if determinism is true and ponders what this could mean. The suggestion Fischer puts forward is that external factors are not just causal contributors of the outcome: they are causally *sufficient* conditions.²³⁵ Now, Fischer grants this distinction between necessary and sufficient conditions. An action can be determined by sufficient conditions in the sense that if certain conditions are in place, it is sufficient for them to cause an event. In contrast, in Fischer's cases which emphasize causal factors like not being struck by lightning, agents lack control over a necessary condition. A necessary condition posits the *failure to occur* of an event that, if it happened, would prevent the possibility of the agent's action.²³⁶ Fischer does not see the importance of the distinction or why it should make a difference. Fischer writes,

In a typical case in which an agent lacks control of a necessary condition of the relevant behavior and yet is subject to robust assessments based on the behavior, it is tempting to say that the condition in question... does not get in the way of or "crowd out" the features of the agent in virtue of which he is morally responsible for his behavior. I am inclined to say that the presence of causally sufficient conditions in the actual sequence similarly need not get in the way of or crowd out the features in virtue of which an agent can be (robustly) morally responsible for his behavior.²³⁷

This distinction between necessary and sufficient conditions is worthy of more discussion than he gives in his essay. The distinction between them is actually crucial because if Fischer is willing to grant that an agent need not be in control of necessary conditions, it is a separate question whether they need to be in control of sufficient conditions.

²³⁵ Fischer, "Sourcehood: Playing The Cards That Are Dealt You," 176.

²³⁶ Fischer, "Sourcehood: Playing The Cards That Are Dealt You," 176.

²³⁷ Fischer, Sourcehood: Playing The Cards That Are Dealt You," 177.

Consider that for some actions it is possible that our personal identity will be a sufficient condition for our behavior. Luther had a sufficient reason to do what he did and to the extent that Luther was able to shape that sufficient reason, his character, he is responsible in the first dimensional sense. The second dimensional sense of responsibility is concerned with responsibility in the first dimension but also for directly shaping the will that is expressed. There will here too be sufficient reasons and conditions for our wills and what this sense of responsibility would seem to require is that agents are themselves among the sufficient reasons and responsible in a significant way for the sufficient reasons being what they are. This will again mean that an agent is not responsible for everything but only something, like her will. If an agent does not play a role in shaping any of the sufficient conditions, then they leave no room for features of our agency which deliver moral responsibility. We can still be morally responsible if we are referencing the first dimension of responsibility, but in the second dimension of responsibility this is not the case. To be responsible in the second dimension of responsibility means that in addition to being responsible for the expression of one's will in action, we have to be responsible for the will itself. Being responsible for having the will in this second dimension sense is understood in terms of being accountable for becoming who we are and what we do. To be accountable requires having the opportunity to shape our ourselves differently in accordance with how we could have had we not been determined.

Fischer and Smilansky focus on the idea of luck. Smilansky's view takes issue with compatibilism because, if determinism is true, there is a sense in which who we are and what we do is just a matter of luck. While we act out of necessity, the sufficient conditions that necessitate them, like our characters, are just lucky. We couldn't have chosen our parents, genes, the time and place we were born. All of these are just lucky and are highly constitutive of who we

become. It is unfair by Smilansky's lights to say someone is fully morally responsible for their behavior when it is produced by things that are a complete matter of luck. If anyone had been born to different parents than the actual ones they had, they would have behaved differently. The arbitrariness of this conclusion is what makes Smilansky dismissive of compatibilism. This is how I interpret his comment about the 'arbitrarily given.' Fischer's response is that the origins of our actions are irrelevant. Focusing on where factors like our characters comes from is as relevant as not being struck by lightning.

As I am offering a sourcehood incompatibilist libertarian view I am sympathetic to Smilansky's concerns about constitutive luck. I also understand Fischer's point that sourcehood incompatibilism can invite an inflated view of agency such that one might think ourselves infinite movers immune from external forces of any kind. My own view of ultimate sourcehood tries to find a middle ground. Ultimate sourcehood need not imply sole sourcehood. It is worth meditating on ultimacy specifically here. Fischer says that no one objects to being the source of one's behavior but to demand that we are an ultimate source is where we run into trouble.²³⁸ Ultimacy is concerned with the past, but we need not only restrict our scope to the past and instead we can also look to the future. Ultimacy is defined in terms of being responsible for a sufficient reason which invites looking into the past to see if it was possible to be responsible for sufficient reasons. Yet we can turn the other way by considering that if we are not ultimate sources, then we only have one possible future which is necessitated by the past. This is inconsistent with being responsible for our wills in the sense required by ultimacy because an agent will never have the opportunity to shape his will differently than he actually will. Being ultimately free means that the future is not set in the way that it is if determinism is true. If

²³⁸ Fischer, "Sourcehood: Playing The Cards That Are Dealt You," 164.

determinism is true, there is only one possible future which necessarily follows from the past. Ultimate free will implies there are different possible futures, and I am the ultimate source of which of my possible futures is the one that is actual because I am able to choose to have a different future than the one I end up having. Ultimacy has a backtracking component since it raises concerns about the past but it is also forward looking because what it is concerned with is agents facing a closed future. For myself at least, this is the most troubling aspect of determinism and a key tension between the ultimacy I think is part of free will and the truth of determinism.

A compatibilist might say that part of what entails the future is the choices we make but an incompatibilist concerned with ultimacy can rejoin that the future that actually obtains is necessary. It cannot be different than what it will be and so it is metaphysically closed. This entails that human agents are not ultimate sources of their lives and identities because what they will become is not contingent upon the alternative possible choices they will make but instead is necessitated by the past and the laws of nature. The future might include the choices I make but the necessity of them entailing from the past makes agents not the ultimate sources of their futures.

If we are not ultimate sources but instead determined, there is a legitimate concern about our responsibility in the second dimensional sense because who we are will just come down to luck. Ultimacy, whether it be regarding sourcehood or responsibility, is concerned about our identities and the behaviors that form them. Ultimate responsibility involves responsibility of the first dimensional kind but more besides. In attempting to listen to her parents and also act contrary to her parent's imperative, Rosa does what she does voluntarily, intentionally, and for reasons and has a limited form of first dimensional responsibility. Her responsibility in the

second dimensional sense is not yet on the scene. It comes about when, after enough times, she has made a choice to set her will, making her the ultimate source of it in this respect.

In summary, the discussion of this chapter has articulated the details of ultimate responsibility and sourcehood. I have argued that for a will to be free agents must be the ultimate source of their wills. In virtue of being the ultimate source of their wills agents are ultimately responsible for their wills.

Chapter 5

Agency and Authorship

1. Authors and Stories

Marya Schechtman offers the following observation: “If the person sitting next to you on a long plane trip suddenly launches into the story of his life you may be amused, or annoyed, or simply glad for the distraction. Whatever your reaction, you are unlikely to be *surprised* that he has a story to tell.”²³⁹ (author’s italics) The film *Forrest Gump* begins in just this sort of way with the protagonist, Forrest Gump, telling his life story to multiple people while he waits for a bus.^{240, 241} In the personal identity literature, there is a view of persons which claims that the stories of our lives constitute our identities. It is called the, “narrative view of the self.”²⁴² However, there is no single narrative account but multiple overlapping views with divergences between them.²⁴³ On the sort of theory offered by theorists like Alasdair MacIntyre, the lives of persons have the form of a narrative; on views like that developed by David Velleman, the nature

²³⁹ Marya Schechtman, “The Narrative Self,” In *The Oxford Handbook of the Self* eds Shaun Gallagher, (Oxford: Oxford University Press, 2011): 396.

²⁴⁰ Robert Zemeckis, “Forest Gump,” Paramount, 1994.

²⁴¹ Marya Schechtman, “Art Imitating Life Imitating Art: Literary Narrative and Autobiographical Narrative,” in *The Philosophy of Autobiography* (Chicago: University of Chicago Press 2016): 22.

²⁴² Schechtman, “The Narrative Self,” 399.

²⁴³ Schechtman, “The Narrative Self,” 399.

of value is narrative; and Daniel Dennett believes we create a fictional self through the stories we tell about ourselves.²⁴⁴

In chapter two I called my own view of free will and personal identity “narrative self-formation” and put the narrative half on hold. I now intend to hopefully reward the reader’s patience by articulating my complete theory. As a helpful summary, the theory I seek to develop in this chapter is the following: fully free will requires that agents are the ultimate source of who they are and what they do. Agents are possessors of this ultimacy only if there are certain undetermined choices in their lives where they form or shape their wills. By the standards of this view, the undetermined choices through which we shape our wills should be contextualized within the broader context of one’s life narrative. By analysing self-forming choices in the context of the story of one’s life, undetermined decisions can be seen as intelligible. The main conclusion I want to defend is that agents are the ultimate source of who they are and what they do only if and only if they are the authors of who they are and what they do. If we were not authors, we would lack ultimate responsibility for who we are and what we do. In this chapter, I want to substantiate the idea of authorship that I claim is necessary and sufficient for free will and ultimate responsibility.

This chapter will explore in detail the relationship between personal identity and free will. A key feature of the view will be to articulate the particular sense of identity that is tied to free will. Furthermore this particular sense of identity is tied to free will through the idea of UR articulated in chapter four. The particular sense of identity that is related to free will is the

²⁴⁴ Schechtman, “The Narrative Self,” 401.; David Velleman, “The Self as Narrator,” In *Self to Self: Selected Essays*. (Cambridge: Cambridge University Press, 2006).; Daniel Dennett, “The Self as a Center of Narrative Gravity,” In *Self and Consciousness: Multiple Perspectives*, eds Frank S. Kessel, Pamela M. Colde, and Dale L. Johnson, (Hillsdale, New Jersey: Lawrence Erlbaum Associates 1992); Alasdair MacIntyre, *After Virtue: A Study in Moral Theory* (Notre Dame, Indiana: University of Notre Dame Press, 1981).

“characterization” sense of identity which is focused on the personal characteristics which make us the kind of person that we are.²⁴⁵ The relationship succinctly put is that persons are UR for their identities through (metaphorically) authoring their own life stories and their character within that story. Being the author of your life story is necessary and sufficient for being ultimately responsible for who you are.

I am going to discuss elements of narrative personal identity common to all accounts which I will utilize in the construction of my own view. My account of narrative identity draws out the relationship between authorship and free will. Specifically, I focus on the role persons play in shaping their personal identities and separate those characteristics which are passive and those which we actively shape. To be sure, our identities are quite complex and come in various degrees of passivity and activeness. I do not claim who we are is entirely self-made but instead I theorize that we are more like a sculptor shaping a pot than a viewer of a movie watching their lives pass them by. To avoid misunderstanding there will still be a fair bit of passivity here too; what one even decides to sculpt could just be passively received while the molding and shaping of it is more active.

Then I move on to discuss what I take to be a serious challenge to my narrative account of identity: the problem of radical personal change or as Schechtman calls it the “problem of multiplicity” The way this presents itself as a challenge to my view is that I suppose that there is unity to the stories that we author. But the possibility of radical personal change suggests that persons are fragmented and not unified. I begin with Laurie Paul and Edna Ullmann-Margalit’s

²⁴⁵Marya Schechtman *The Constitution of Selves*, (Ithaca, NY: Cornell University Press, 1996).

work on decision theory.²⁴⁶ Decision theory studies what a rational procedure for deliberating between any set of options should be.²⁴⁷ Both Paul and Ullman-Margalit have argued that the standard model of practical rationality is insufficient for describing how agents rationally decide to acquire new values.²⁴⁸ They both discuss moments where one faces a big decision which will alter oneself and presumably one's values.²⁴⁹ I develop an answer to this problem for my model of free will and authorship. I enlist Agnes Callard's account of aspirational agency to aid my view. Her model of agency has many important tools that I believe we can exploit to propose a persuasive response. I will say more when I come to it, but a brief statement of the reply is that understanding identity as a fluid construction within a narrative structure can show how the problem of radical change is manageable.²⁵⁰

2. The Nature of Narratives

I begin with proposing and developing the nature of a narrative. Noel Carroll offers an account of narrative explanation that strengthens the explanatory power of narrative.²⁵¹ Carroll's thesis is to define a crucial aspect of narrative, what he calls the 'narrative connection'.²⁵² In doing so he elaborates other features definitive of narrative. He says that narratives are

²⁴⁶ Laurie Paul, *Transformative Experience*. 1st ed. (Oxford: Oxford University Press, 2014); Edna Ullman-Margalit, "Big Decisions: Opting, Converting, Drifting." *Royal Institute of Philosophy Supplement* 58 (2006): 157–72.

²⁴⁷ Agnes Callard, *Aspiration: The Agency of Becoming*, (New York: Oxford University Press, 2018): 39.

²⁴⁸ Paul, *Transformative Experience*, 17; Ullman-Margalit, "Big Decisions: Opting, Converting, Drifting," 164.

²⁴⁹ Ullman-Margalit, "Big Decisions: Opting, Converting, Drifting," 160.

²⁵⁰ Marya Schechtman, *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. First edition. (Oxford: Oxford University Press, 2014).

²⁵¹ Noel Carroll, "On the Narrative Connection," in *Beyond Aesthetics: Philosophical Essays* (New York: Cambridge University Press, 2001).

²⁵² Carroll, "On the Narrative Connection," 118.

comprised of more than one event/state of affairs in a globally forward-looking manner pertaining to the career of at least one unified subject. A narrative must also have temporal relations between the events and/or states of affairs in a perspicuously ordered manner, and finally the earlier events in the sequence must be at least causally necessary conditions for the causation of later events and/or states of affairs.²⁵³

To anticipate an objection, I suspect readers might think of stories which don't fit these criteria as counterexamples. Carroll helpfully groups different *story forms* into three categories: *annals*, *chronicles*, and *narratives*.²⁵⁴ Thus, Carroll's thesis does not make the obviously false claim that stories which are not narratives do not count as stories. Indeed, Carroll is inquiring into what the difference is between these story forms and what the characteristics of narrative are. Annals and chronicles are stories but they are not narratives. Some examples will help. "The Eastern Roman Empire falls in 1453; the American Constitution is accepted in 1789; Russia is defeated in 1905."²⁵⁵ This is an annal. An annal is a story form since it has a temporal ordering to it. However, it only has temporal ordering, i.e., one thing happens after another with no connection to previous events. "The French Revolution occurred in 1789; Napoleon became Emperor in 1805; Napoleon was defeated at Waterloo in 1815, after which the Bourbons were restored."²⁵⁶ This is a chronicle because it has a unified subject, French history, and a perspicuous temporal structure. Carroll argues that chronicles are not narratives because of the lack of a connection between events other than that of temporal succession. According to Carroll,

²⁵³ Carroll, "On the Narrative Connection," 126.

²⁵⁴ Carroll, "On the Narrative Connection," 121.

²⁵⁵ Carroll, "On the Narrative Connection," 121.

²⁵⁶ Carroll, "On the Narrative Connection," 121.

“If I say ‘I woke up; later I dressed’ still later I went to class,”” I suspect that most people would agree that this falls short of a full-fledged narrative, although the events cited might be turned into ingredients of a narrative.²⁵⁷ But why isn’t it a narrative properly so called? To put it vaguely- because the connection among the events alluded to by it is not tight enough.”²⁵⁸ What makes the connection ‘tight enough’? Carroll answers that the connection is a causal one. For example, “Creon had Antigone executed; consequently, his son committed suicide, which caused his wife to commit suicide, and, as a result, Creon felt anguish.” This is a narrative.²⁵⁹ It contains a unified subject, Creon, it is perspicuously time-ordered and it is causally linked. The previous events in the discursive string are the causes of later events in the string, “... in the sense that the earlier events supply sufficient grounds, all things being equal, for the occurrence of the alter event.”²⁶⁰ I think Carroll is on to something and so I also believe that narratives are typically and temporally structured. The example just given of Creon illustrates the causal and temporal aspects of a narrative.

Carroll’s causal claim goes on to say that causation is important for narrative but it needs qualification. He does not advance deterministic causation as a demarcating feature. According to Carroll, most narratives are not strings of causal entailments as earlier events underdetermine later events. For example, “the thief enters the bank to rob it, but subsequently, as he exits, he is apprehended by the police” is a narrative where the first event is not the fully determining cause

²⁵⁷ This is a contentious point where readers are likely to have different intuitions from Carroll.

²⁵⁸ Carroll, “On the Narrative Connection,” 122.

²⁵⁹ Carroll, “On the Narrative Connection,” 122.

²⁶⁰ Carroll, “On the Narrative Connection,” 122.

of the second event.²⁶¹ There are some events in narratives that might work this way but it is not true in every case. Carroll offers the following argument about the causal relations in a narrative:

Of course, in saying that the earlier event in a narrative connection is *at least* a causally necessary condition... one leaves open the possibility that sometimes the relation may be stronger causally. Sometimes the earlier event may be the cause of a subsequent event and/or state of affairs in the sense of necessitation, or sometimes a series of earlier events and/or states of affairs may be jointly sufficient for the production of a later event. I do not preclude these possibilities in some cases. Rather, I deny that they supply an account across the board of the narrative connection. A better candidate for that account is that the earlier event and/or state of affairs in a narrative connection is at least a causally necessary condition or ingredient for bringing about later events...²⁶²

The events and/or states of affairs in a narrative operate in a causal network but the causal relations are held together by necessary causal conditions and some sufficient ones. If determinism is true every event is causally necessitated by earlier events. Sufficiency is also related to determinism in that if a set of conditions is sufficient for an event to occur, and those conditions are not met, even if they are necessary, the event will not occur. Whether those conditions are or are not met is determined by prior factors which are necessary for the sufficient condition to be met. If there is only one sufficient condition to play tennis and it is my motivation and I have no motivation then I do not play tennis. The motivation to play tennis is necessary to want to play tennis while actually having that motivation is sufficient to play tennis. Without the sufficient condition in play the event of my playing tennis will not occur.

I think though we should recognize that there are non-causal events which we bring about that are also temporally structured. If my thesis supervisor breaks his promise to read my thesis when he said he would, the not-reading constitutes breaking the promise. The causal aspect would be what caused my supervisor to not read the thesis while the act of not reading the thesis

²⁶¹ Carroll, "On the Narrative Connection," 122-123.

²⁶² Carroll, "On the Narrative Connection," 125.

is constitutive of the promise breaking. Suppose as revenge for his promise-breaking, I cut his brake lines and thereby put him in danger. The act of cutting the brake lines is constitutive of putting him in danger. I put him in danger just by cutting his brakes.

This would seem to open the possibility of narratives that do not involve causality at all. Interestingly, the non-causal case retains the temporal structure of narrative: to put someone in danger is forward-directed while promise breaking is backward-looking. While these actions themselves have no causal relations, it seems open to Carroll to say that they could come to play them. As Carroll says, something which by itself does not fit into his view of narrative could be an ingredient in making a narrative. When I cut my thesis supervisor's brake lines and thereby put him in danger, the endangering is non-causally related to the cutting, but it could be an ingredient into a narrative with causal relations. "Andrew got into an accident because his brake lines did not work and his grad student felt regret for cutting the brake lines upon hearing of the advisor's accident." The earlier non-causal act of endangering Andrew by cutting his brakes comes to take on a causal role in the later event of the car accident and the grad student's emotions. We can therefore, I think, make a modification to Carroll's view where non-causal events can form narratives but causal events are far more typical and important

I think it is worth focusing in on exactly what Carroll is after in his discussion of narrative. He wants an account of the "narrative connection" which I take to be the nature of the relations between events which allow them to be described as a coherent whole narrative. The alternative models of story-telling contain events which might be put together to tell a story, but they don't fit into a single whole narrative. Narratives are distinguishable by virtue of having causal and temporal relations even though they can sometimes have non-causal ones.

3. Applying Narratives To Persons

When we think about the life of a person, there will obviously be events which include things they do causally and events with things they do non-causally. It also includes episodes which have nothing to do with each other such as things like eating oatmeal for breakfast and eating chicken for dinner. A proper account of persons in terms of narratives will be one where persons have lives with a narrative structure. It should focus on how certain features and particular events can relate to each other in terms of their broader place in the structure of a narrative. This allows for an account of narrative to consistently incorporate causal and non-causal relationships between events.

This is consistent because to hold that narrative involves causation when describing the life of a person, it seems unnecessary to require every event in the life of a person have a causal relation to each other. There are far too many obvious counterexamples of things persons could do which are local and unrelated to each other. But it does not follow that there are no events which fall into a narrative structure with the appropriate causal and temporal relationships.

In the life story of persons, there are prior events which play a causally necessary and sufficient role for later ones. However, prior events will sometimes only play a causally necessary role and not always a sufficient one. I postulate that in the narrative of our own lives, these moments will be decisions where one's character, powers, states of mind and the like are not sufficient for later events. I further surmise that at least some of these events with necessary but no sufficient causes will be self-forming actions. In SFAs, factors like character and motives and state of mind influence and incline us in some ways rather than others. For example, the businesswoman is torn between helping and going to her meeting, but she never once entertains joining in and participating in the mugging. Our explanation for her action can use something like her character to explain why she didn't do that. Her character, however, is not sufficient for

her to have a motive strong enough to make her decisive in choosing to help the victim or go to her meeting. The details of her prior life story do not provide a sufficient cause. But from the details of her life story and character up to that point, it makes sense to a metaphorical reader why she is torn.²⁶³ The businesswoman when making her SFA has reasons which satisfy for her decision either way she makes. A typical feature of SFAs is that they are made for satisficing reasons and not sufficient reasons.

Consider the following case introduced by Bernard Williams of the French painter Paul Gauguin deliberating between going to Tahiti and staying in France.²⁶⁴ Williams describes Gauguin as being in a dilemma as he wants to pursue his ideal of becoming a great artist but he also does not want to abandon his wife and children. The details of Gauguin's life story provide causal conditions that make the two options the only rational ones. Gauguin's personal commitments to his values and his ambitions to become a great artist are the necessary conditions that provide inputs into his deliberation. They do not fully explain his choice but provide direct connection between his past and whatever choice he makes. The prior events of Gauguin's life generate satisficing reasons as he has reasons that are good enough whatever he chooses. In other words, it makes sense from Gauguin's perspective to do whatever he decides to do of the two possible alternatives.

If Gauguin's narrative generated sufficient reasons for him to pursue his artistic ambitions, it would be irrational or unintentional from his perspective to stay. If he were to do otherwise it would not be what he wanted to do. On the other hand, if his narrative generates

²⁶³ Kane, "The Problem of Free Will: A Libertarian Perspective," 65.

²⁶⁴ Bernard Williams, *Moral Luck: Philosophical Papers 1973-1980* (Cambridge: Cambridge University Press, 1981).

satisficing reasons, either possible alternative could be done on purpose and also rationally because neither has sufficient weight to rule out the other. This moment in Gauguin's (metaphorical) story is not confined to only one possible choice but to two possible choices. It is possible for the prior events of narratives to generate satisficing reasons for multiple alternatives. I think this part of what it is to think of the lives of persons as narratives is important because it recognizes that our lives can be, though need not, complex and can develop in lots of different ways. To think of our lives as a narrative does not imply that our lives have a single end point or that we must aim at one single goal. As we get further into the discussion, we will see this last point is a common criticism of applying the metaphor of a narrative to the lives of persons. But first we must lay the groundwork to understand what it means to describe persons as having a life story.

4. Narrative Views of Personal Identity

Let's discuss narrative views of personal identity. In the literature, this theory of persons is often called narrative views of the self. They are well-known in contemporary discussions of personal identity. The first of two types of narrative view I consider is the hermeneutical variety. This view is well known from the work of Charles Taylor, Alasdair MacIntyre, and Paul Ricoeur.²⁶⁵ While it should be noted there are differences in the way each of them understand the lives of persons to be a narrative, they all agree that persons are fundamentally "self-interpreting beings."²⁶⁶ The starting point for hermeneutical theorists is the claim that a person is inherently

²⁶⁵ MacIntyre, *After Virtue*; Paul Ricoeur, *Oneself as Another* (Chicago: University of Chicago Press, 1992).; Charles Taylor, *Sources of the Self: The Making of the Modern Identity* (Cambridge, Mass: Harvard University Press, 1989).

²⁶⁶ Schechtman, "The Narrative Self," 398.; Ricoeur, *Oneself as Another*.; Taylor, *Sources of the Self*.; MacIntyre, *After Virtue*.

a narrative entity.²⁶⁷ There are two further claims attached this thesis: (1) that a person's sense of themselves must be narrative, and (2) that the lives they lead have a narrative structure.

The distinctive feature of this approach to a narrative view of the self is the ties all three thinkers draw between persons, narrative, and agency. On this view, agency requires an intelligibility of our actions to ourselves and others.²⁶⁸ Consider, as an example, the question "Should I have taken the job offer and moved away from my friends and family?" According to the hermeneutical theorists, we can only intelligently understand our behavior with an interpretive lens directed at our narratives. Reflecting on whether I ought to have pursued the career opportunity at the cost of distancing myself from familial ties, I interpret this present decision in light of how my past shaped and indeed lead to it. If I have been very close to my family up to that point, moving far away would be a cost to pursue the career I have spent time cultivating.

MacIntyre emphasizes the significance of interpreting an agent's intentions through referencing the context of their life story.²⁶⁹ We could describe an agent doing the same action in different ways and each description would have "equal truth and appropriateness," for example, digging, gardening, exercise, preparing for winter, or simply pleasing one's wife.²⁷⁰ The way to understand which description is the correct one of the agent's behavior is the one that captures his intention. We can ask "why are you gardening" and the answer to this will make reference what matters to the agent in the action they are performing and this will mean understanding the

²⁶⁷ Schechtman, "The Narrative Self," 398.

²⁶⁸ MacIntyre, *After Virtue*, 204.; Taylor, *Sources of the Self*, 56.; Ricoeur, *One Self As Another*, 159-161.

²⁶⁹ MacIntyre, *After Virtue*, 204; 206; 208.

²⁷⁰ MacIntyre, *After Virtue*, 208.

action's causal, non-causal, and temporal relations. (1984: 208). When we do this, we will end up with a narrative, and thus, "narrative history of a certain kind turns out to be the basic and essential genre for the characterization of human action"²⁷¹ Demarcating action from mere occurrence requires we be able to, "identify it under a type of description which allows us to see that occurrence as flowing intelligibly from a human agent's intentions, motives, passions, and purposes."²⁷² What I think MacIntyre means here is that the project of action demarcation needs an understanding of what end or goal I am aiming at. A separate but related part of this project is examining origins, but not every action will have deep or interesting histories. Scratching an itch is a goal directed activity and its origins are uninteresting. I scratched because I was itchy and that is all. Sometimes, however, it is necessary to know how an intention, plan, motive or purpose develops and this development will require taking a broader view. My accepting or rejecting a job offer would be an action but its intelligibility would be through the broader context of the decision in my narrative.

Another important aspect of the hermeneutical view is that our stories be aimed at some end, goal or, as MacIntyre puts it, *telos*. According to MacIntyre, for our lives to be intelligible we cannot just put them into stories; these stories have to be ordered towards some further ethical end²⁷³. To lead a good life for MacIntyre is to search for and aim for the good. The unity of a human life," MacIntyre says, "is the unity of a narrative quest."²⁷⁴ I think this brings out one of the key issues with narrative personal identity: the issue of the unity of a life. As I do intend to

²⁷¹ MacIntyre, *After Virtue*, 208.

²⁷² MacIntyre, *After Virtue*, 219.

²⁷³ MacIntyre, *After Virtue*, 208.

²⁷⁴ MacIntyre, *After Virtue*, 208.

propose my own view of narrative unity, I put this issue on hold for now while I get clear on the hermeneutical articulation of narrative.

Another important feature of hermeneutical theories of narrativity is the association between persons, selves, and agency. An example of the ties between these concepts is found in the work of Anthony Rudd. Rudd says,

To think of who I am in narrative terms is to think of myself as a temporal being and so as a being with a past; but also to understand my current situation as the result of my past history, not just causally, but in the sense that my past history has established the meaning of my present situation. Moreover, it involves thinking of myself as an agent (rather than a mere passive spectator of life) and therefore as someone who is partly responsible for the shape his life has taken. And it is also to understand myself as a being with an (indeterminate) future, which I am attempting to shape in certain ways that are intelligible to me (and at least potentially to others) as arising out of or responding to my current situation and the past that has shaped it. I take it that normal human agents do have such a sense of themselves as characters in a developing story, though it will usually be largely implicit and unarticulated.²⁷⁵

The first-person perspective described in this passage describes the conditions hermeneuticists believe are necessary for a narrative sense of self and agency. It is useful to recall Carroll's account here: on his view a narrative involves events which are connected by causal relations and temporal orderings. There are causal relations in Rudd's description of the first-person perspective. The past stands in a causal relation to the present and future as the past shapes the meaning of the present and the possibilities for my future that I am attempting to shape. Note: there are things in the past which might shape the present but we weren't aware of but it doesn't appear necessary to be aware of all the things in the past for the point that the present is influenced by the past.

²⁷⁵ Anthony Rudd, "Kierkegaard, MacIntyre and Narrative Unity—Reply to Lippitt" *Inquiry*, 50, no. 5, (2007): 543.

MacIntyre seems to also stress the temporal element of Carroll's narrative view but goes beyond causal relations. His account aims at capturing the intentions of an agent when they act and as I have shown, which he tries to do so by contextualizing them in a temporal way. I think here too we could understand that in evaluating my options they don't just pop up from nowhere on MacIntyre's account, they instead come from a past that has developed in a certain way. It is plausible to say that there are still causal relations here, but that MacIntyre sees the evaluation from the agent's perspective as important to explain an action. The reason it seems to me this matters to MacIntyre is his goal of separating action from mere occurrences. Actions are things we do on purpose while mere occurrences don't have intentions but still have causes and effects. An account of the explanation of action has the agent and their intentions at centre stage as it shows that actions are not just a string of causes and effects but that they are brought about by agents for reasons that matter to them.

The present discussion of a narrative account of persons results in persons having a narrative structure that has both causal and non-causal relations as well as temporal relations plus an evaluative component of the present. The narrative structure of a person's life might not always lead to persons thinking of their lives as stories, but we can point to cases where people are aware of their present and future sensibly emerging from their pasts. Consider cases involving persons suffering with dementia and their loved ones. When people go to see their loved one, the pain they might feel is a result of knowing who that person was in the past is diminished and that in the future there will be even less remaining of that person. My Italian father lives with dementia and this is how it feels for not just myself but for my siblings too; the man with whom I watched calcio and ate pasta when I was a kid is at least not metaphorically the same person I

visit at the nursing home, and I know in the future he will one day not recognize me. This story-like structure of my cognition is how it seems our lives are like stories.

An issue for the hermeneutical variation of narrative identity is its strong sense of agency which excludes individuals that by normal standards would be viewed as persons. Consider as an example an agent I will call Charles: Charles has to deliberate between whether he goes to culinary school in Rome or to the local dentistry school in Chicago. The hermeneutical account of narrative intertwines Charles' agency with his personhood in the following way: Charles does not merely deliberate between the two options: he thinks about his action options explicitly, but his understanding of these options is at least implicitly shaped by knowledge and/or understanding of these that is least implicitly shaped by past, which and is also sometimes the focus of his explicit conscious thought.

That is a tall order for Charles to meet in order to say that he is an agent. There are groups of beings who do not meet this condition that it would be, I think, just false not to include under the concept of agency or persons. To take another personal example, my brother is disabled with autism spectrum disorder and he certainly cannot act or think in the way the hermeneutical account requires to count him as an agent or a person, but this is problematic. In one sense agency is just an ability or power to pursue and fulfill one's desires. This is the sense I identified in the accounts of classical compatibilism in the previous chapter. My brother certainly does this and much more besides; he constructs his whole daily routine, he makes his own breakfast, and is able to communicate sufficiently about what he wants. The assertion that human beings like my brother are not to be seen as either agents in any sense or persons is implausible if not straightforwardly false.

Despite this, I think that there are pieces of the hermeneutical approach to narrative identity that are worth keeping. I think it is possible to articulate a narrative view without an inflated notion of agency. What I want to keep is the view of our lives as having narrative structure or the very least a story-like structure. This allows us to have a more inclusive concept of agency while retaining a narrative theory of personal identity.

5. Narrative Lives

A key feature of a narrative approach to personhood is that the lives of persons have a narrative structure. We need not think about ourselves as a character in a story in the same way we think of literary characters and narratives. Rather, narrative is a structure that arranges our lives as we develop as persons. The developmental child psychologist Katherine Nelson suggests that persons use narratives in their development from infancy to maturity writing, “the hypothesis of a new level of consciousness that emerges in early childhood together with a new sense of self situated in time and in multiple social realities.”²⁷⁶ This is based on her observation of the gradually complex kinds of self-awareness that are seen in the normal development of a human infant. She sees the emergence of a “new subjective level of conscious awareness, with a sense of a specific past and awareness of a possible future, as well as with new insight into the consciousness of other people.”²⁷⁷ Nelson links this level of consciousness to the development of the ability to tell simple narratives about one’s life, and so she finds a “close connection between narrative and the emergence of a specifically human level of consciousness.”²⁷⁸ Nelson believes

²⁷⁶ Katherine Nelson “Narrative and the Emergence of a Consciousness of Self.” In *Narrative and Consciousness*, (New York: Oxford University Press, 2003): 17.

²⁷⁷ Nelson “Narrative and the Emergence of a Consciousness of Self,” 33.

²⁷⁸ Nelson “Narrative and the Emergence of a Consciousness of Self,” 22.

that as children enter into language, with the help of caregivers, they learn to narrate events in their lives. This begins with describing things that happened to them in sequence and offering basic evaluations (e.g., “Mom and I went to the park and fed the ducks. They were cute. It was fun.”) By Carroll’s standards this isn’t a narrative, but we need to only point out they are only just starting to learn how to tell a story. With help from adults, they also begin to tell not just stories but proper narratives with causal relations with temporal orderings. Nelson describes how initially children need prompting from adults to do this but eventually they do it by themselves, and when they do they experience qualitatively new kind of experiences. She says, “This level of self-understanding integrates action and consciousness into a whole self, and establishes a self-history as unique to the self, differentiated from others’ experiential histories...” It adds “a new awareness of self in past and future experiences and the contrast of that self to others’ narratives of their past and future experiences.”²⁷⁹

While the focus so far has been individual my narrative view of persons also has a relational component. I think this is an important point to make because authorship ought to include the ways in which who we are is shaped and influenced by our roles and relationships with other people. To the extent that we are so influenced, we have to recognize that other people help to write our stories as co-authors. In early work, Hilde Lindemann investigates the ways that oppressive narratives can damage identity and how the construction of an alternative narrative can repair that identity.²⁸⁰ She defines identity in narrative terms:

An identity is a representation of a self; it’s a narrative understanding of who someone is. It is generated from both an internal and an external perspective, and consists of the tissue of stories and story fragments that are woven around the acts, experiences, personal

²⁷⁹ Nelson, “Narrative and the Emergence of a Consciousness of Self,” 7.

²⁸⁰ Hilde Lindemann, *Damaged Identities, Narrative Repair*, (Ithaca: Cornell University Press, 2001): 6.

characteristics, roles, relationships, and commitments that matter most about a person—either to her or to others around her”.²⁸¹

What is useful about this definition of narrative identity is that it has internal and external components. Narratives are written by an individual, but these stories not (metaphorically) written in isolation but alongside other people. Schechtman utilizes the social component of this view of narrative to offer a view of narrative personal identity that allows for infants and persons with cognitive defects to be persons. While these people cannot narrate their own lives the way others can, their identities are created by narratives written by other people. Schechtman clarifies that this claim is, “not about the way in which others can form narrative conceptions of what infants and the demented are *like* (although of course they can do that as well) but rather about the way in which they can begin or continue an *individual* life, bringing past and future into the present, as it were, on behalf of the person who cannot do it herself.”²⁸² The interactions between caregivers and infants are an example of this. As Schechtman notes we

talk to infants and very young children as if they have a level of self-consciousness or narrative skill that they do not yet have, and we treat them as moral and prudential beings before they have become such.... Eventually, if all goes well, they will become a competent giver and receiver of justifications. We make two-year olds apologize when they have taken someone else’s toy or hit someone not because we think they are fully developed moral agents who should be regretful for what they have done, but because we are trying to teach them to be such agents.²⁸³

From these two perspectives, we can see how our lives are like narratives and in a way that is broad enough to include people that the hermeneutical theory would exclude. To anticipate a possible worry, one might ponder if these remarks lead one to making everything persons. Since

²⁸¹ Hilde Lindemann, “Holding on to Edmund: The Relational Work of Identity.” In *Naturalized Bioethics: Toward Responsible Knowing and Practice*. eds Hilde Lindemann,, Marian Verkerk, and Margaret Urban Walker. (Cambridge: Cambridge University Press, 2009): 72.

²⁸² Marya Schechtman, *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. First edition. Oxford: Oxford University Press, 2014): 104.

²⁸³ Schechtman, *Staying Alive*, 104.

I am comparing the lives of persons to stories it might seem that anything which can be the subject of a story is given personhood. I think the answer is clearly negative. We can see this from both angles as infants are in the process of not being authors of their own lives but with others practice becoming the narrator or author of her own story. From the other side, dementia patients are in the opposite case of having had the ability to author their stories but now they cannot. My father used to be able to tell me all about his life in Italy and now he cannot. Both infants and dementia patients presently lack authorship but from a unified perspective retain their personhood because they will become authors in the case of infants or authored their lives up to a point and no longer can in the case of dementia patients. Things which can be said to not be persons noncontroversially like rocks and fridges never have this ability to author themselves or a unity they authored. In cases like the earlier case of my brother with autism he is plausible said to be in a perpetual state like the infant in the sense of needing assistance in authoring his life.

The first-and-third-person perspectives discussed provide an understanding of how to understand our lives as a single unified narrative. Obviously, real life is not as tidy or rigid as literary narratives or sonatas, but just because our lives are less neat it does not follow that there is not a structure to them. It is worth emphasizing that we have a structure not necessarily a goal when talking about narratives and persons. The discussion has not tried to establish personhood in terms of a single goal that a narrative is constructed around, and this has been no accident. This might appear contrary to the importance of intentions—what a person aims at—but I think this concern can be handled by seeing intentions guiding and guided by narrative structures. Intentions guide narratives in the sense that some of the things we aim at will shape the structure. Persons do not intend explicitly to shape their narratives but rather the things they aim at will do that work. The intentions themselves are guided by narratives in the sense that what an agent will

aim at will be shaped by the necessary, sufficient, and satisficing reasons generated by their narratives. Recall the earlier example of Gauguin going to Tahiti to become a great painter. His intentions are shaped by his life story in that the considerations that motivate him to act are what they are because of the shape of his narrative. The choice Gauguin makes will shape his life story, but he does not choose to do what he does thinking about trying to create a single life story about a great French painter. Gauguin's intentions are clearly guiding and guided by the story of his life and this understanding shows that intentions do play an important role but without requiring us to say he is aiming at one unifying goal.

What narrative views of personal identity are committed to is the present point of a narrative structure. To have a structure will imply a certain kind of unity but it is a unity of the life of a person and not an end, goal or purpose. Nelson's observation of how young children develop shows we don't think of ourselves as living a story towards a single goal; instead, having a life narrative is about a structure that organizes events in a particular way. This organization brings things together in a unified way which can be seen from the third-person perspective as well. When I think about my father when I was young and he was healthy and think about him now with dementia, the judgements I make are about the same person even though he is radically different from how he was. To think about him as the same person, I am thinking about a narrative structure that unifies a subject and allows for judgements and descriptions of one person as the same person over time.

It is useful to tie this point back to the basic view of narratives: I will focus on the causal and temporal aspects of narratives. I have also said they include an intelligibility of what matters to persons. The causal relations and temporal orderings of events are linked to the present point as they help bolster the emphasis of a unified structure. There are earlier events which have a

causal relation and temporally order later ones and which shape or influence what matters to us. This won't be true of everything that matters at every moment as we might come to value something new that is not part of the past but could influence later events. An important part of a narrative structure tied to the present conception of narrative is that there are things that really matter to us which are extended through time that help to define who we are. Consider, as an example, Jane who has a bake shop. Jane has had many failed attempts at starting her own business but this time her bake shop is successful. With every venture, the particular episode is causally related to the past and her understanding of each failed attempt fits into her narrative as a whole. "That last attempt failed so let's try something different" with every new store opened until her latest kick at the can works. The various attempts might seem unrelated to each other, and in one sense they are, but in another significant sense they fit into her narrative as a whole. "Finally," she thinks, "all my hard work has finally paid off after years of failed attempts. All of my labouring has not been wasted and my goal of running my own business is now a reality." The meaning of the success is temporally structured and causally related to her past and her present beliefs about her future. Her sense of her present accomplishment and what she can and should expect in the future is only intelligible because of her past failures.

This brings out the first piece of my view of persons in terms of authorship which is what Schechtman calls the characterization sense of identity. Schechtman defines this sense of identity in terms of those "actions, experiences, beliefs, values, desires, character traits, and so on," which can be attributed to an individual person.²⁸⁴ This sense of identity is focused on attributability of attributes to agents. While our identities are complex, I am interested in those personal characteristics we strongly identify with and feel make us who we are. If I think being

²⁸⁴ Schechtman, *The Constitution of Selves*, 73.

kind is an integral part of the kind of person I am this an attribute attributable to me and picks me out as a certain kind of person. Of course, there will be characteristics which we wish to be rid of as well as passive ones which also are part of the kind of person we are. There is a case to be made that what we identify as important to us is partly constitutive of who we are. The personal characteristics which are thought to be inseparable from identifying a particular person. Furthermore, I think it can be shown that there is at least one sense in which such an identification is narratively structured and only intelligible from the perspective of the unified whole. That is to say we think about the personal characteristics which we identify as integral to the characterization sense of identity and there are times where this identification will be narratively structured.

Consider the following example to help illustrate this point. In the film *Stranger Than Fiction*, the protagonist Harold Crick comes to learn that he is actually a character in a novel by an author named Kay in her latest work *Death and Taxes*.²⁸⁵ Harold learns that Kay is plotting his demise as she does with all of her protagonists. We have no background on Harold, and it seems fruitless to speculate that he had a life prior to what we the viewer see. Harold exemplifies the necessity of narrative for self-reflective identification. Schechtman makes a distinction between thinking one's life is a narrative and taking on what she calls a narrative attitude.²⁸⁶ One believes one's life is a narrative in the straightforward sense that one is a character in a novel being written by an author for the purposes of dissemination and consumption. The narrative attitude is essentially taking a perspective on one's life from the perspective of a narrative. What this entails is that one thinks about one's life as a unified whole but need not do so at every

²⁸⁵ Marc Forester, *Stranger than Fiction*. Culver City, California: Sony Pictures Home Entertainment, 2007.

²⁸⁶ Marya Schechtman, "Art Imitating Life Imitating Art: Literary Narrative and Autobiographical Narrative." In *The Philosophy of Autobiography*, eds Christopher Cowley, (Chicago: University of Chicago Press, 2015): 28.

moment. To take on this attitude would only require there be moments where we do. Harold serves as an illustration of this point as there is one specific moment in the film where he does exactly that. When Harold enters a music shop and picks up a guitar, he reflects on himself and his life as a whole and he finds it lacking. Alongside the knowledge that his life is a story, he can see that he does not like the kind of story he is living in, how it has developed and how his life would continue if he did not do anything about it. For he sees that his life is missing something he actually does care about: music. If and only if Harold takes this perspective of a whole life is he able to recognize dissatisfaction with his life and that he wants to change it. This seems to me to be typical of a wide range of persons. Now, in those cases such as dementia patients or infants, it is the people helping to write the narrative which take a narratively structured reflective view of the person. Despite the obvious difference in behavior, I think of my father as the same person in a unified way even though in the present moment he is not the same as he was in the past or will be in the future.

This account is different than the one found in MacIntyre as I am not saying we are on an ethical quest. Instead, we are shaping the narrative structure of our life stories and give our life stories meaning; narratives serve persons and not the other way around. In this way, the account of unity acknowledges that there are important differences between life and literature, as what may be good for our lives might not be good story-telling. But they are not wholly independent in terms of shaping the lives of the characters which in real life are us. Harold demonstrates both of these points well as his death is necessary for the external author (Kay) to write a good story and this is obviously bad for Harold. Yet it would be a mistake for Harold to give up taking a literary perspective where he examines his life as whole and to return to his previous cold-calculating life that he judged as needing changing.

6. Authoring Narratives

This brings me to the role of free will in the construction of our narratives and being the authors of them. *Stranger than Fiction* deliberately plays around with the extent to which Harold is an actual author with control and the source of his decisions. Harold definitely feels and experiences his life as free and as if he is the author of it. But it is not clear that the motives and desires that are part of Harold's identity are shaped by him. This is where the issue about free will begins to enter the discussion of authorship. What is the basis of Harold choosing to shape his life? Is it Kay or is it Harold? Does Harold choose to be who he is of his own free will?

As I think of it, free will involves the ability to author the story of one's life and the kind of person that the story is about. This conception of free will ties into the characterization sense of identity through Schechtman's contention that characterization identity has four features that are of concern when we ask questions about personal identity. These four features are self-interested concern, survival, compensation, and moral responsibility.²⁸⁷ I will elaborate on moral responsibility's connection to personal identity and then discuss its relevance to my theory.

The nature of moral responsibility is a contentious matter and so to avoid any question-begging definitions I am focusing on moral responsibility in the sense of being deserving of praise or blame for one's actions. This leaves moral responsibility open to involving more if it indeed does involve more than just praise and blame.

The association between moral responsibility and personal identity is in what counts as 'one's actions,' i.e. what actions are really mine. The connecting thread between moral responsibility and personal identity is in whether a morally praiseworthy or blameworthy action

²⁸⁷ Schechtman, *Constitution of Selves*, 3;73.

as well as omissions can be truly attributable to a person. The scope of the present discussion is about responsibility for one's own actions and omissions.

I want to tie personal identity in the sense of the characterization question to moral responsibility through Kane's notion of UR. Specifically, I draw a tie between identity and the ultimacy clause of UR. I claim that the U clause gives us an answer to how to attribute moral responsibility to an individual. Satisfying the ultimacy clause requires that an agent is responsible for at least a sufficient cause, reason or motive. Moreover, it doesn't mean in every instance they do something that makes them responsible for the sufficient cause, reason or motive. Rather, responsibility is derivative from actions which did not have sufficient reasons. These would be self-forming actions and performing SFAs makes one responsible for some of the sufficient reasons when making a non-SFA. By making an SFA, an agent is the ultimate source of her action since she is responsible for some of the sufficient reasons.

If being morally responsible for one's actions requires being the ultimate source, what kind of free will is necessary to make it the case that I am the ultimate source of my identity and my action? A kind of free will consisting in the ability to author one's identity. I think this understanding of free will connects back to the distinction Kane raised between different senses of freedom when we talk about free action and free will.²⁸⁸ When discussing the problem of free will it is easy to slip into framing the problem as whether freedom is compatible with determinism. This is a mistake because a term like "freedom" does not have one univocal use. Isaiah Berlin famously identified and distinguished at least two different kinds of freedom.²⁸⁹

²⁸⁸ Kane, *A Contemporary Introduction to Free Will*, 6.

²⁸⁹ Berlin, "Two Concepts of Liberty".

Just as Berlin demarcated between ‘negative’ and ‘positive’ freedom, we need to sort out the kind of freedom at issue in free will debates.

I believe that if we think of free will in terms of being the authors of our life stories, we will satisfy all the conditions needed to be ultimately responsible and the ultimate source of who we are. Authorship entails that persons are actively creating themselves and freely create themselves into the kind of person that they are. Authorship provides the ownership that is necessary to attribute an action as belonging to me. Here is a rough sketch of my argument.

1. Moral responsibility for one’s own actions (and omissions) requires that one have an ownership over and for one’s will.
2. A necessary condition for ownership for my will is that it is attributable to me.
3. For my will to be attributable to me I must be the ultimate source of it.
4. In order to be the ultimate source of my will I must have authored it and authorship requires that one satisfies the ultimacy condition of ultimate sourcehood
5. This is not possible in a deterministic universe.
6. Therefore, being an author of one’s will is incompatible with determinism.
7. Therefore, either we are the author of our wills and determinism is false or we are not the authors of our wills and determinism is true.
8. We are the authors of our wills.
9. Thus determinism is false.

The first premise makes the claim that moral responsibility requires ownership and the second premise says ownership requires attributability. There are passive and active senses of ownership and we need to keep them separate. If there is anything we can say a person owns it is their body and this kind of ownership would be passive in the sense that bodies genuinely belong to us simply because we were born into them. There is nothing you have to do to keep ownership of your body. We might come to make adjustments to our bodies like tattoos and so forth but what you do with what you own doesn’t change the relationship of ownership itself. If we consider for the sake of argument, that something like a job is something you have to keep doing for it to be yours. . There is this active sense of ownership in moral responsibility in the differentiation between accidents and intentional action. The first premise can be read in terms of the

identification aspect of ownership because we can say an agent is morally responsible for who they are when they take ownership of. This seems to me to be exactly the move made by compatibilists like Frankfurt and Watson. Their views rely on an identification relationship between the actions of persons and their wills which is the basis for an agent to have ownership for their action.

The sense of attributability in the second premise is ambiguous as it is possible to mean just attributability—I am the owner of my will if it is indeed mine—which would be uninteresting and unclear. As with ownership, there are active and passive senses of attribution. We can again use our bodies as an example as the grounds to attribute my body to me are passive. My body just is mine in virtue of the relationship between me and it.²⁹⁰ The active sense of attributability that is more interesting is like that of a literal author. This thesis is attributable to me on the grounds that I wrote it. Moreover, it is because of the decisions and choices I made that it has its structure and shape. This links to the third premise, that my will is attributable to me because I authored it and we have a more clear sense of the intended meaning of attributability. Keeping passive and active senses straight is important because we would have a false sense of attributability that denied a passive sort of attributability. An active sense is important though because there are clear cases where attributability depends on things we do. Persons who plagiarize are not properly attributable as authors of the work or ideas they claimed to have authored. This provides a bridge between ownership in the first premise to the third one.

The first two premises then lead to an intermediate conclusion that satisfying U of UR requires that I have a particular sense of ownership for my will and that I am the owner because

²⁹⁰ Conjoined twin cases might be interesting cases where it is unclear how we carve out the attribution. Is it one body attributable to two people or one person with half a body each or one person with an atypical body? I take no stance but these cases are acknowledged here as they are noteworthy anomalies.

certain aspects of my will are attributable to me. We then have a further intermediate conclusion that ultimacy is necessary and sufficient for authorship. In other words, it is necessary and sufficient to be authors that we are the ultimate source of our wills. The rest of the argument is an empirical point and so I stop at the seventh premise because the truth of determinism is not philosophical but scientific and an open question at this point in time. So, the focus of the chapter and the defence of my argument needs to be focused on authorship. I have discussed the literature on perspectives of personal identity that rely on the metaphor of a narrative and discussed the idea of narrative itself. To demonstrate what I take to be an author in the sense of having free will, we need to understand authorship. What does it mean to be an author of an autobiographical narrative if we assume that our lives can be intelligibly thought of in that way? This is the next task to embark on our descent down incompatibilist mountain.

7. Narrative Self-Formation

The most obvious example of authorship is people who actually write stories. They reflect on what they want their stories to be and get on with the task of writing them. An author deliberates whether they want their story to be dramatic, poetic, or adventurous and how to develop the characters in line with the kind of story they want to create. At some point they will need to examine the work as a whole. It might be as they work through a draft or it could even be after writing an initial first draft, but in order to correctly align their story with what they want it to be, they will have to look at it in totality to make those judgments about it. By analogy, people as self-authors would be persons who construct their identities and live in similar ways.

My view of authorship says that there is an important way persons do this through the choices they make. In what other way could we understand authorship? Perhaps we could understand it in terms of the conditions necessary to be an author. Literal authors need to be able

to write; this means they need the ability and the resources to create a piece of writing. Metaphorical authors would also need these things, but how do we understand the ability and resources for metaphorical writing? Since my theory is an incompatibilist view, one condition for metaphorical authorship would be indeterminism. In what way does indeterminism function as a condition for authorship? It is not my contention that indeterminism per se is sufficient for authorship. As I said in chapter three with Austin's putting case, that a choice is undetermined is necessary for a choice to be free, but it is not enough; we must also satisfy other conditions. We can't just have the mere ability to do otherwise but we have to be able to do otherwise on purpose and rationally. This is because in situations where agents can satisfy certain conditions like being able to have done otherwise voluntarily, intentionally, rationally and on purpose, agents are able to genuinely author their life stories and characters. This in turn enables agents to satisfy another, and perhaps more important, criterion for authorship: ultimate sourcehood. Only in circumstances where agents are not determined to act and have robust alternatives are their decisions "up to them." They are not *the sole reason* for why they do whatever they choose to do but they are the *ultimate* or "buck stopping" reason.

I argue this reasoning leads us to conclude that authorship in the metaphorical sense I am interested in involves being able to do or perform specific actions. These actions would seem to fit the description of Kane's self-forming actions. In SFAs, agents have robust alternatives as the dilemma they are in is only generated by the robust nature of the alternative possibilities. SFAs also satisfy the view of narrative, as the prior events of an agent's story provides the influences which incline without necessitating an agent. An agent making an SFA also acts on the basis of satisficing reasons. I earlier discussed Gauguin's dilemma between staying in France and going to Tahiti. Gauguin's situation is a good example of both my view of authorship and self-

formation. Gauguin as I see it is reflecting on what he wants his life to be and what kind of life he wants to live. In deliberating over this, he is taking a perspective of his life as a whole, as he could only decide what kind of life he wants to live by examining the one he has now from a unified perspective. We can see that there is a narrative structure to Gauguin's life because of the way that the present possibilities for his future have been influenced and shaped by his past. To decide which of the possibilities to pursue requires he thinks about himself in a unified way.

We can author ourselves without explicitly thinking about ourselves as characters in a story. I don't think such a strong sense of narrative is required or likely true. When making an SFA, we can see that the way the narrative structure of a person has developed is a major reason for why they are in that situation. If the businesswoman was completely indifferent to the mugging, she wouldn't even hesitate to ignore the mugging victim and just go on about her day. The person she is at that point has developed from her past and the possibilities for the future (albeit the immediate future) in the present are all related to each other. This point obtains without the businesswoman thinking she is a character in a narrative. So, in the act of making an SFA and authoring her story, she is shaping the structure of her story without necessarily thinking about herself as a unified character writing a story. There might be situations in which thinking of oneself in this way is necessary. Gauguin might be such a case. The key point is that there is a unifying narrative structure more so than conceiving of one's life as a unified narrative story. The latter would be to take an attitude that one is living in a story while the former is about what it is to have the life of a person.

Return to SFAs. A person facing a SFA does not have a clear vision of how their narratives should continue based on the character or the will of their character. The personal history of agents provides causal inputs which incline agents but because they are not necessitating

(deterministic) their futures remain open. The arc of the narrative is shaped by how it has unfolded but the present and future are not necessitated by the past so how the story continues is up to the author, i.e., the agent. Being the author of one's life story is to have the freedom to choose to shape one's will and one's life how one wants to. This understanding of free will is incompatible with determinism because if we are truly the authors of our wills we are the ultimate sources of them and ultimacy in this sense is incompatible with determinism. We thus have a case for the incompatibility of both free will and moral responsibility with determinism through authorship and ultimate sourcehood.

An often-neglected aspect of SFAs that Kane has touched on in his writings is that every SFA is also the initiation of what Kane calls a value experiment.²⁹¹ Value experiments are characterized by Kane as decisions whose justifications are in the future and are not fully explainable by an agent's past. Kane says that romantic dates, marriages, career choices, economic and political policies are all examples of value experiments.²⁹² Each of these ordinary cases involves risk and an uncertainty over the success or failure of the outcome. The nature of such experiments is that they are open to verification through testing any expectations against experience. Value experiments can also be plans of action or ways of life and can be found in many parts of our ordinary every day lives.

²⁹¹ Kane *Free Will and Values*; Kane *Significance of Free Will*, Kane, "Libertarianism", Kane, "The Dialectic of Selfhood," Kane, "Rethinking Free Will," Robert Kane, "Acting 'of One's Own Free Will': Modern Reflections on an Ancient Philosophical Problem." *Proceedings of the Aristotelian Society* 114 (2014).

²⁹² Robert Kane, *Ethics and the Quest for Wisdom* (Cambridge: Cambridge University Press, 2010).

Kane's Kierkegaardian inspiration for value experiments is seen in that every value experiment may lead to despair.²⁹³ The inspiration from Mill is seen in that an underlying motivation about value experimentation is that some truths about the good life cannot be learned from the armchair or a priori but only through practical engagement with the world. Since every value experiment might have disappointing results. A person facing a SFA does not have a clear way to sort out which ones will be successful is to live them out. A qualification of this point is that we can actually engage in value experiments in our mind. Kane says that value experiments can and often, "do take place in the head – imagining scenarios, considering possible outcomes and consequences of actions in the attempt to decide which plans or intentions are worthy of being pursued in reality."²⁹⁴ The importance of testing them out through living them is that we are fallible. We need something to test our hypotheses against lest we run the risk of spinning fanciful tales in our heads.

The value experimentation within SFAs takes place within a larger narrative structure, the story of one's life. This is true in virtue of how one's past has shaped one's present. SFAs are events that occur when our previously formed character and motives are conflicted in the present moment. The options before me are partially circumstantial but they also are related to my character. It is not as simple as either just circumstance or just character. It is circumstantial as I am in a situation that requires I make a choice but it is also related to character because the options before me are meaningful in virtue of who I am and who I want to become, as well as other things. We make our characters and identities from the situations we find ourselves in. We weave together our various experiences via value experiments initiated in SFAs to create a life

²⁹³ Soren Kierkegaard, *Fear and Trembling and The Sickness Unto Death*, trans. by Walter Lowie. (New York: Anchor Books, 1954); Mill, *On Liberty*.

²⁹⁴ Kane, *Ethics and the Quest for Wisdom*, 84.

that is good for us. SFAs and therefore value experiments are to be explained within a wider narrative context. This is because the conflict distinctive of self-formation arises from how our life stories have unfolded at that moment. The explanation of value experiments is one which conveys the intelligibility of the choice made within the wider narrative context of one's life story.

While this might seem to be in tension with the emphasis on history my view of identity has, this is misleading. Value experiments are wider than just SFAs so one might engage in a value experiment without performing an SFA. But every SFA is a value experiment and in this context agents make value experiments because of a lack of decisive reason ruling one way rather than another. Since they are in a dilemma, they run a value experiment and decide what to do. The businesswoman has to decide which set of reasons she will be moved to act by and then she will come to see if her expectations meet the actual outcome. Another example would be marriage. There are two separate questions to answer through value experimentation. There is the initial decision of whether one should get married in the first place and then the separate question of whether it will turn out to be the life-long relationship one hopes it to be at the beginning. We can see through these aspects of value experimentation how value experiments tie into diachronic unity and personal history. For one's past provides causal input into the decision; one's experiences over time provide influence and guidance into shaping the present circumstances and one's present feelings. The expectations one might have about the future are also seen as being related to the present which has been shaped by or is seen as a continuation of the present.

8. The Problem of Multiplicity and Change

There is a worry about the basic framework I have outlined that I am calling narrative self-formation. This is a problem I am labelling the problem of radical change. My theory implies that there is a sustained diachronic identity across time as my view of identity casts personal identity as a set of characteristics which are attributable to persons. These characteristics are attributable to individuals in a way that allows us to track them over time. If Jones is kind and I can continuously attribute that characteristic to Jones over time, then there is a sense in which Jones is the same person over time.

The problem of radical change poses a worry for my theory because it undermines the existence of a diachronic conception of identity. I begin with Laurie Paul and Edna Ullmann-Margalit's work on decision theory.²⁹⁵ Decision theory studies what a rational procedure for deliberating between any set of options should be.²⁹⁶ Both Paul and Ullman-Margalit have argued that the standard model of practical rationality is insufficient for describing how agents rationally decide to acquire new values.²⁹⁷ They both discuss moments where one faces a big decision which will alter oneself and presumably one's values. Ullman-Margalit calls these decisions "big decisions" and she and Paul believe there is a problem with decision theory in it cannot tell us how to make such decisions rationally. As they see it, the problem is that,

In order to make a rational decision to ϕ (for the sake of A) instead of ψ -ing (for the sake of B), one needs to know whether (and by how much) one prefers A to B and whether (and by how much) it is more likely that ϕ -ing produces A than that ψ -ing produces B . The problem is that because one (or both) of the options promises a substantial change in preferences, the agent doesn't have a single, stable set of preferences that could provide the input for the decision procedure. Over a period of time during which my desires are unstable, there may be no fact of the matter about which option would better fulfill them.

²⁹⁵ Paul, *Transformative Experience*; Ullman-Margalit, "Big Decisions: Opting, Converting, Drifting".

²⁹⁶ Callard, *Aspiration*, 39.

²⁹⁷ Paul, *Transformative Experience*, 17.; Ullman-Margalit, "Big Decisions: Opting, Converting, Drifting," 164.

Decision theory assumes relatively stable preferences, and an agent making a big choice conforms a radical shift in her preferences.²⁹⁸

Ullman-Margalit discusses the problem in terms of the rationality of big decisions from a third-person perspective. She puts the problem like this:

New Person is now, by hypothesis, a transformed person. Opting transforms the sets of one's core beliefs and desires. A significant personality shift takes place in our opter, a shift that alters his cognitive as well as evaluative systems. New Person's new sets of beliefs and desires may well be internally consistent but the point about the transformation is that inconsistency now exists between New Person's system of beliefs and desires, taken as whole, and Old Person's system taken as a whole. I am not questioning his ability to actually make a choice, or his ability subsequently to assess himself as happy (or unhappy) with his choice. The question I am raising is whether it is possible to assess the rationality of his choice, given that his choice straddles two discontinuous personalities with two different rationality bases"²⁹⁹

Ullman-Margalit refers to making a "big choice" as "opting" and describes it with the following example

I was told of a person who hesitated to have children because he did not want to become the "boring type" that all his friends became after they had children. Finally, he did decide to have a child and, with time, he did adopt the boring characteristics of his parent friends—but he was happy! I suppose second order preferences are crucial to the way we are to make sense of this story. As Old Person, he did not approve of the person he knew he would become if has children: his preferences were to not have New Person's preferences. As New Person, however, not only did he acquire the predicted new set of preferences, he also seems to have approved of himself having them. How are we to assess the question whether he opted "right"?"³⁰⁰

Ullman-Margalit argues the problem for decision theory is that there is no neutral perspective to which we can ascend to judge the rationality of the choice. If we take the perspective of Old Person, the choice made to have a child is a bad one, but for New Person this is the right choice. While Old Person did not want children, New Person is happy to be a parent which is in direct

²⁹⁸ Callard, *Aspiration*, 42.

²⁹⁹ Ullman-Margalit, "Big Decisions: Opting, Converting, Drifting," 167.

³⁰⁰ Ullman-Margalit, "Big Decisions: Opting, Converting, Drifting," n. 10.

opposition to the preferences of Old Person. For Ullman-Margalit there is no way to assess if the choice to have children was the “right one.” To be sure, Ullman-Margalit is not charging us with being irrational with respect to “big decisions”, only that individuals are arational.³⁰¹ To be irrational, there would have to be a rational way of deciding in such cases as “big decisions” and Ullman-Margalit denies there is such a way.

Paul’s approach to the topic is like Ullman-Margalit but she adopts a first-person perspective in her work. She is concerned with the difficulty of accounting for the potential value of novel experiences that come about from what she calls our “epistemic and personal transformations.”³⁰² She has an example that is used throughout the book where the reader is offered the opportunity to become a vampire: “As a member of the undead, your life will be completely different. You’ll experience a range of intense, revelatory new sense experiences. You’ll gain immortal strength, speed and power, and you’ll look fantastic in everything you wear. You’ll also need to drink blood and avoid sunlight.”³⁰³ For Paul there is no way for us mortals to have the knowledge necessary to make the choice to become a vampire in an informed way. While you might very well decide to take the plunge what Paul would insist on is you have not the slightest idea what you are getting yourself into.³⁰⁴ Such personal transformations are like Ullman-Margalit’s “big decisions” and just like Ullman-Margalit, Paul is skeptical of making these decisions in a standard rational way.

³⁰¹ Ullman-Margalit, “Big Decisions: Opting, Converting, Drifting”.

³⁰² Paul, *Transformative Experience*, 17.

³⁰³ Paul, *Transformative Experience*, 1.

³⁰⁴ Paul, *Transformative Experience*, 17.

Paul discusses a dimension of decision making that Ullman-Margalit leaves out in her discussion of “big decisions.” Paul supposes that you could conduct some empirical research about becoming a vampire: you could ask all your friends and other loved ones who became vampires about their experiences. Do they enjoy the nocturnal routine or having to drink blood? Turns out they do!³⁰⁵ You can conduct more thorough research and interview both people who did not become vampires and those who did and see what the net satisfaction is overall. We can also apply this to the previous example of choosing to have a child.

Now it turns out that there have been studies that have been done on this. Specifically, it has been done on the happiness of those with kids and those without them.³⁰⁶ Paul cautiously concludes that the rational option would be not to have children on this empirical approach:

(I)f you are prepared to ignore all of your subjective assessments about what it would be like for you to have a child, and choose solely on the basis of the empirical research, if you want to maximize your expected subjective value, the research (to the extent that there are clear results) suggests you should not have a child. If we simply follow the dominant empirical conclusions of the experts, it seems that anyone who wants to make a rational decision about parenthood that is based on maximizing expected subjective value must either suspend judgment (given the lack of a clear censuses on the results) or should actively choose to remain childless.³⁰⁷

Paul does not believe, however, that one should make a decision based on such studies. She argues that a person who makes a choice based on what she calls “impersonal, “big data” reasoning” sacrifices her autonomy.³⁰⁸ She writes,

Unless robots have taken over the world when humans weren’t looking, for many of us this is an untenable way to approach choices involving our personal goals, hopes, projects, and dreams. In other words, in today’s society, when making important personal

³⁰⁵ Paul, *Transformative Experience*, 1.

³⁰⁶ Callard, *Aspiration*, 45.

³⁰⁷ Paul, *Transformative Experience*, 87.

³⁰⁸ Paul, *Transformative Experience*, 88.

choices, we want to consult our wants, personal preferences and to reflect on what we want our future lives to be like as part of assigning values to outcomes. It is simply unacceptable to be expected to give up this sort of personal autonomy in order to make decisions about how one wants to live one's life.³⁰⁹

Her point seems to be that one should consult one's preferences first and foremost and not outsource one's decision making to uninterpreted and impersonal data. When we are deciding such matters as should I marry, if so then whom, what career do I pursue and so forth, we have to reflect on our preferences and judge what this will mean for me. Ullman-Margalit and Paul deny that one can rationally know what a future career will be like for me because me as I am now a person with different values than from the person in the future will become.

Ullman-Margalit and Paul illustrate the problem of radical change because their work discusses precisely the problem of diachronic identity. Their work as I have presented it believes that there is no diachronic identity in terms of values. Both of the authors have argued that the persons we become are radically different from the persons that we are and there is no way we can say there is continuity between these people. We simply change from one kind of person with a set of values to another and we become new people all at once. This poses a direct challenge to my view because my view of narrative self-formation says that there is a diachronic structure which has a unity to it. My view of identity says we can describe ourselves over time through different distinct changes and that the changes to our identity in the specified sense of characterization happen through our choices. Ullman-Margalit and Paul's challenge to my view is that there is no diachronic personhood or rather there is reason to doubt a unity of persons with values that can be attributed to them over time. We are certain people are at one time with certain values and then just radically become different people with different values and no continuity

³⁰⁹ Paul, *Transformative Experience*, 87.

between the transition. I need to provide an answer to this worry, and I have to show how we can rationally choose to become who we are.

9. Authorship and Identity

Agnes Callard's work on agency is helpful here. Thinking about agency in terms of her concept of aspiration helps to show how agents are both passive and active in the formation of their identities; Callard's model of aspirational agency also lends itself to building a plausible account of authorship (though she herself does not have anything to say about narrative views of personal identity).

Callard engages with the criticisms raised in the last section of the standard decision model in developing her theory of aspiration. These criticisms are at a very high level of generality as neither are concerned with a particular model of decision-making or rationality. Ullman-Margalit characterizes a rational decision as seen by the decision-model as, "the best way of satisfying the full set of the person's desires, given his or her beliefs formed on the basis of the (optimal amount of) evidence at their disposal."³¹⁰ Paul argues that a rational decision maker ought to, "choose the act that has the highest expected value."³¹¹

To take the parenthood case Ullman-Margalit raised earlier, there is a shared assumption about the best way of analyzing rationality. We should take a person who is currently a non-parent and examine their decision to become a parent. This is where we will see the rationality of this decision. Callard's model of aspirational practical rationality argues that one can indeed be rational in going through a personal transformation even though one will acquire new values and

³¹⁰ Ullman-Margalit, "Big Decisions: Opting, Converting, Drifting," 164.

³¹¹ Paul, *Transformative Experience*, 21.

change who they are. Her complaint with both thinkers is that their generalizing approach casts all decisions as being made in the same way. She challenges the assumption that rational deliberations are an in the moment all or nothing phenomenon. In her view a person can be rational in say a procedural way rather than a structural way. This is part of her project to define her concept of “aspiration”.

Here is Callard’s characterization of aspiration: “the rational process by which we work to care about (or love, or value, or desire...) something new is ‘aspiration.’ Aspiration, as I understand it, is the distinctive form of agency directed at the acquisition of values.”³¹² Callard’s project is to provide “a philosophical account of how it is possible for value-acquisition to be a form of practically rational agency”³¹³ Callard believes the difficulty with this project is exactly the one outlined by Paul and Ullman-Margalit. Callard says:

People do not seem to be able to *choose* or *decide* to have different values. A decision or choice is an act of the will that prefigures, accompanies, or is a constitutive part of some single action. The transition from indifference to love cannot typically be effected by way of deciding: no matter the strength of my will, it does not seem that I can muscle myself into suddenly caring. To be sure, the path to valuing sometimes includes momentary expressions of commitment: the moment when you say “I do,” or sign the adoption papers, or buy the one-way plane ticket to a foreign country. But these moments are themselves only part of the story, punction of a longer process. Coming to value something tends to represent a deep change in how one sees and feels and thinks. Acquiring a new value often alters the structure of one’s priorities by demoting or even displacing something one valued before.³¹⁴

So then how do we change and become different people with different values? On the rare occasion we can change radically and quickly. Phineas Gage is one such case of radical and quick change: due to an accident Gage had his skull impaled by a railway spike and somehow

³¹² Callard, *Aspiration*, 4.

³¹³ Callard, *Aspiration*, 2.

³¹⁴ Callard, *Aspiration*, 2.

survived. According to reports from Gage's friends, he was no longer Gage.³¹⁵ Gage is an interesting case but a rare one and so it is a case to keep in mind, but I am going to focus on more typical gradual changes in personality. We should not misunderstand gradual change as not counting as radical. Change that is slow can be just as radical and, in some cases, even more so. I think that in typical cases, changes in our values happen somewhat passively but are also a product of what we intentionally do. Typically, becoming a new person is not just something that happens to us but neither do we just radically transform into someone new. We come to be someone piece by piece. Callard argues, "I think that gradual transformation of preference can be one way in which a person's agency is manifest; not only is becoming something little by little something that happens to us, it is also, sometimes, something we do..."³¹⁶ Moreover,

Unlike becoming a vampire, becoming a parent is neither something that just happens to you nor something you decide to have happen to you. It is something you do. And it is not something you can do in a moment: we spend a long time becoming mothers and fathers. When one makes a radical life change, one does not submit oneself to be changed by some transformative event or object; one's agency runs all the way through to the endpoint. The nature of that agency... is one of learning: coming to acquire the value means learning to see the world in a new way.³¹⁷

Of course, there is the obvious sense in which someone just becomes a parent; when a child is born. But Callard's point employs the sense of "parent" in terms of personal characteristics and not biology. It is the sense of "parent" that a foster, step, or adopted child would use as the adult in their lives don't just become their parents. They are legal guardians but in the eyes of the child the adult has to work at becoming their parent. From the adult's perspective, they have to work at becoming the parent of the new child and that typically results in adopting the values of that person which they currently are not. The process of coming to acquire values and adopting new

³¹⁵ The actual true extent of his change remains unknown.

³¹⁶ Callard, *Aspiration*, 57.

³¹⁷ Callard, *Aspiration*, 61.

perspectives is what Callard means by aspiration. The aspirant currently does not have the values they want to hold; to aspire to those values means they learn about them as they work towards acquiring them.

In her view, a person who is not currently a mother aspires to becoming a mother and is in the process of becoming a mother. In the aspirational model, we actively but gradually become what we are via aspirational agency. The aspirant is one who comes to learn what it is to be what they want to become and starts to actually become that person. One learns gradually as one inches closer to being what they aspire towards. This is not in Callard's view something that just happens to us. Learning about what it would be like to be a certain kind of person with certain values is less like reading a book and more like exercising. We do not just imagine what it would be like to become a certain kind of person; we work towards it.

Callard's account of aspirational agency has more dimensions, however, the materials necessary for my purposes are at hand so we needn't go any further. Her account of aspiration has, I think, the building blocks for a rejoinder to the original worry I raised about radical transformations. Callard's account of aspiration describes the process of becoming who we are as something we do rather than merely something that happens to us. Of course, there are other factors at play not under our control that shape us, but such factors are alongside the active process of identity-construction. Her account of aspiration shares with my view a relationship between personhood and agency, and both of our accounts cast this relation in terms of personal values. The issue of radical change on Callard's account is shown to be not as serious as it appears.

While radical change might appear to challenge any kind of narrative unity to personal identity, Callard's view aids in showing this not to be the case. The reason is that there is a

rational process by which we change, and this process includes our agency. Even if we change radically, the point remains that we have a hand in shaping who we become. There are exceptions to be sure such as Phineas Gage but given the details of such cases they are so extreme as to be rare. Aside from these outliers, we have a way of defending the view that our identities have a narrative unity. Generally speaking, aspirants reflect upon who they are now and who they want to be and to go from one to the other requires taking the necessary steps to actually be that person; this is the point of her discussion about parenthood. Aspirants are aware of their pasts to the extent it is relevant to shaping their present lives and the expectations they have about the future. The form of agency that aspiration takes has a structure to it that fits with a narrative conception of identity. Since this form of agency is aimed at how we change our values even radically, this provides a solution to the concerns raised by Paul and Ullman-Margalit. As aspirants we change ourselves at least partly through things we do and what we even aspire to is related to what we already value. To value something new may lead to displacing already existing values but the process by which this happens fits within a narrative structure.

Here is an example of Callard linking aspiration to the performance of actions:

If one aspires to be a doctor, one goes to medical school. If one aspires to be more attuned to values of healthy living, one might become a member of a gym and transition one's eating habits toward vegetables and whole grains. If one seeks to appreciate some person, one might invite him for coffee. If one aspires to be religious, one might spend more time at one's church or synagogue or mosque—or, in another kind of case, one might deliberately stay away from those places in an effort to (re)connect with God on one's own terms. If one's goal is to value civic engagement, one might explore community activism. We aspire by doing things, and the things we do change us so that we are able to do the same things, or things of that kind, better and better.³¹⁸

From the perspective of my own narrative theory, we think about what we presently aspire towards in the future and this is shaped by our pasts. If a person aspires to reconnect with their

³¹⁸ Callard, *Aspiration*, 5.

personal faith, they would have to have recognized that they lost touch with their faith. On Callard's account of valuing something, I think there is more to be said regarding aspiration and the acquisition of new values. If I aspire to be a doctor I do indeed have to graduate from medical school, and this could result in acquiring new values. On Callard's account, new values might displace existing values and this brings us to an important point. Imagine Sarah who aspires to be a doctor and so has to go to medical school. It is quite far away from home and so going would mean losing relationships she really values. It is reasonable that she could become conflicted over whether to indeed follow through with this aspiration or to stay in her home town. Regarding her personal values it is plausible to say that she could be conflicted over whether to pursue the new values of becoming a doctor or to hold on to her present values. Her present situation illustrates how even if she goes on to radically change, the process by which she does fits within the framework of narrative self-formation. Her radical change was something she did and it was made with reference to her present expectations about the future being shaped by her past.

Acquiring new values runs the risk of displacing existing values and surely sometimes we would be aware of this possibility. However, the process of becoming a doctor and acquiring those new values would typically have, at least, an initial moment of reflection.

With an account of unity in relation to aspiration articulated I can now offer an answer to the problem of radical change. Consider again becoming a parent. My account of agency and persons says that we choose to become who we are through shaping our identities and calls this process authorship. Imagine Mary aspires towards being a parent and along the way has doubts about continuing to work at becoming this person. She makes SFAs in these moments of doubt and doing so she continues to work at acquiring the values to be that person. Mary is the author

of her identity because she chose to continue to shape it this way and it was open to her to continue or to refrain from sustaining her goal. Her life has a narrative structure, and she shapes it in these moments when it is open to her to continue or refrain from becoming the kind of person at the end of the process. Each of us is both an author of and character in our own life stories, making ourselves from our pasts which do not limit us to one future pathway.

In summary, this chapter has proposed, articulated, and defended a view of free will and personal identity. I have argued for what I have called the narrative self-formation view (NSFV). I put forward and defended a narrative view of personal identity where an agent has a single unified narrative. I described how agents are the authors of their own life stories, and that being an author is linked to being an agent. I defended the NSFV from the objection that our lives are not unified and so we lack a unified narrative. I argued that the unity of our stories comes from the choices agents make to shape their narratives and the manner in which agents make these choices demonstrate narrative unity. This addresses the narrative side of the NSFV. On the self-formation side of the view an objection that looms large comes from the problem of luck. I now turn my attention to this issue.

Chapter 6

Libertarianism and Luck

1. The Luck Objection

This chapter engages with an important worry for accounts of free agency that require indeterminism for freedom.³¹⁹ Al Mele presents a specific formulation of the problem as the “problem of present indeterministic luck.”³²⁰ Since I am going to focus on just this problem, I shall call it the problem of luck. The problem put briefly is that if an agent’s choice is undetermined, there is no feature of an agent which can explain why an agent made one choice rather than another.³²¹ Without this explanation, it is alleged that an agent does not control her action because it is just a matter of luck. If an action is not within our control and just a matter of luck, the agent is not the ultimate source of it. Consider Mele’s example of Joe: “In the actual world, Joe decides at *t* to *A*. In another world with the same laws of nature and the same past, he decides at *t* not to *A*. If there is nothing about Joe’s powers, capacities, state of mind, moral character, and the like in either world that accounts for this difference, then the difference seems

³¹⁹ For instance, Derk Pereboom argues that event-causal libertarianism is subject to a kind of luck objection he calls the “disappearing agent objection” which argues that if events are explained in terms of states and events the agent plays no role in producing action. Whatever we might say about this objection and others in the family of the problem it is unmanageable to deal with all of them. I focus on one of them because it is the most pertinent and I believe my solution in this context can be utilized to address these other problems at a later date.

³²⁰ Alfred Mele, *Free Will and Luck*, (New York: Oxford University Press, 2006): 201.

³²¹ Alfred Mele, “Kane, Luck, and the Significance of Free Will.” *Philosophical Explorations* 2, no. 2 (1999): 96–104.; Mele, *Free Will and Luck*.

to be just a matter of luck.”³²² The problem is why do agents act as they do given they could have done otherwise. We seem unable to explain why identical agents make divergent choices. It is clear that Mele’s challenge threatens the very ultimacy at the heart of the sort of libertarianism I have proposed.

I attempt to mitigate the luck objection through articulating the relationship between agents and their SFAs. I argue that this relationship is the narrative structure or story of one’s life and that the features of one’s narrative structure preserve ultimacy. I am going to demonstrate this while granting that the difference in cases like Joe is luck. By demonstrating that the ultimacy my view maintains to be critical for free will is compatible with luck, the force of the luck objection is disarmed. Even though luck is why an identical agent in cross-worlds makes a different decision, I argue that ultimacy under some conditions still is importantly connected and follows from the agent in such a way as to maintain the status of agents as ultimate sources and ultimately responsible.

2. Luck, Control and Responsibility

I will begin with a discussion of the various meanings of ‘luck’ before outlining Kane’s attempt to solve the problem and Mele’s rejoinder to Kane’s solution. What does it mean to say that an event which occurred or not occurred is a result of luck? ‘Luck’ is an ambiguous term as we can think of different types of cases where luck is involved. Sometimes luck is associated with a positive outcome and sometimes it is associated with a negative outcome. There is also a sense in which we can talk about luck neutrally. An instance of an outcome that is a result of luck and would be positively connotative would be (assuming no cheating was involved)

³²² Mele, *Free Will and Luck*, 9

winning the lottery. There is also the phenomenon of ascribing good luck to oneself or to others. “Jones is gifted with good luck. Good things just seem to happen to him” is a common sort of expression. An instance of an outcome that is a result of luck and would be negatively connotative is being struck by lightning. Just as with positively connotative luck we also ascribe negative lucky outcomes to a personal characteristic. “I naturally have bad luck” “Jones is an unlucky guy”.³²³ There is also a way to think about luck which has no good or bad outcomes. An example of this sort of luck would be brute facts like the law of gravity. It would be quite strange to attribute goodness or badness to things like gravity. We can go deeper than this in our analysis of luck as luck can enter into the production of action. For example, a baseball player hitting a homerun involves both some control by the agent but also some luck. While the agent controls the swing of the bat as well as some of the training involved to develop the necessary skills, there are other factors at play beyond their control, like the strength and direction of the wind, which determine if the agent succeeds in hitting a homerun. An interesting feature of this is that we praise agents if they are successful at achieving their goals even though luck was involved in producing the outcome. Professional athletes are widely celebrated and highly compensated if they perform well at their sport even though luck plays a role in their success.

Morality intersects with luck in quite an interesting way. Immanuel Kant famously claimed that morality is immune or independent from luck:

A good will is not good because of what it effects or accomplishes, not because of its fitness to attain some proposed end, but only because of its volition, that is, it is good in itself... Even if, by a special disfavor of fortune or by the... provision of a step motherly nature, this will should wholly lack the capacity to carry out its purpose—if with its greatest efforts it should yet achieve nothing and only the good will were left (not, of course, as a mere wish

³²³ As an interesting aside, I think an intriguing question for x-phi philosophy would be the propensity to ascribe to oneself good luck versus bad luck. That is to say, are people more inclined to think of themselves as naturally lucky or unlucky? What factors influence their judgements about themselves in this regard?

but as the summoning of all means insofar as they are in our control)—then, like a jewel, it would still shine by itself, as something that has its full worth in itself. Usefulness or fruitlessness can neither add anything to this worth nor take anything away from it.³²⁴

In a pair of now classic papers of the same name Bernard Williams and Thomas Nagel challenged Kant's assertion that morality is independent of luck. They introduced what is now known as the control principle:

The Control Principle (CP): We are morally assessable only to the extent that what we are assessed for depends on factors under our control.³²⁵

This principle is held alongside another intuitive corollary principle:

(CP-Corollary) Two people ought not to be morally assessed differently if the only other differences between them are due to factors beyond their control.³²⁶

We see the plausibility of these two principles by applying them to particular cases. For instance, “if two drivers have taken all precautions, and are abiding by all the rules of the road, and in one case, a dog runs in front of the car and is killed, and not in the other, then, given that the dog's running out was not something over which either driver had control, it seems that we are reluctant to blame one driver more than the other.”³²⁷

Nagel offers four classifications of moral luck: resultant luck, circumstantial luck, constitutive luck, and causal luck.³²⁸ Here they are more formally defined:

Resultant Luck: This is luck in the way things turn out. This applies to immediate actions and long-term goals. Consider again the case of athletes who could have their careers ended at

³²⁴ Immanuel Kant, *Groundwork of the Metaphysics of Morals*, ed. and transl., Mary Gregor, (Cambridge: Cambridge University Press, 1998).

³²⁵ Nelkin, Dana K., "Moral Luck", The Stanford Encyclopedia of Philosophy (Spring 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.), <https://plato.stanford.edu/archives/spr2023/entries/moral-luck/>.

³²⁶ Nelkin, “Moral Luck”.

³²⁷ Nelkin, “Moral Luck”.

³²⁸ Thomas Nagel, *Mortal Questions*. First Edition, (Cambridge: Cambridge University Press 1979).

any point from an injury. Or consider an entrepreneur trying to make money selling their product. The inventor of Zoom surely could have never seen how lucky they could be creating their product just in time for a global pandemic to necessitate the entire world to use their platform resulting in becoming fabulously wealthy.

Circumstantial luck: This is luck in the circumstances in which one finds oneself.³²⁹ Kane's businesswoman example is actually one instance of this as it was beyond the businesswoman's control that a mugging that she would witness would take place that day and on her way to work. Had the businesswoman not been there when the mugging took place, she would have not either helped or ignored the victim. If we imagine a businesswoman who was not there and we judge them the same way, then this would be a case of circumstantial moral luck.

Causal luck: The third category is luck in "how one is determined by antecedent circumstances" (Nagel 1979,60). Nagel believes that this luck is the essence of the problem of free will as he says the actions of agents appear to be the consequences of causally antecedent circumstances beyond the control of agents. A significant feature of causal luck is the that the truth or falsity of determinism is irrelevant because whatever the causal origins of the prior conditions of action agents will have no control over them.

Constitutive Luck: This is the luck of being the kind of person that one is. Genes, upbringing, peers, and other environmental factors influence us and we have no control over these matters so to that extent luck constitutes who we are. If an agent is cowardly, for instance, and this a result of character traits or the kind of person they are, then acting cowardly is caused somewhat by luck.

³²⁹ Nagel, *Mortal Questions*, 9.

There has been a substantial literature around the topic of moral luck and Nagel and Williams' discussion of it. There is another form of luck which is distinct yet could be said to have elements of at least two sorts of luck. Kane says it has been called the "Explanatory Luck Objection", stated in the following way by Alfred Mele:

If different free choices could emerge from [exactly] the same past [as their being undetermined would require] there would seem to be no explanation for why one choice was made rather than another in terms of the total prior character, motives and purposes of the agent. The difference in choice, i.e., why the agent made one choice rather than another, would therefore be just a matter of luck.³³⁰

We can see that the key issue here is that there seems to be no relevant feature of the agent that explains why the agent made one choice rather than another. We might say further still that since there is no factor that settles and causes the behavior, the outcome has an element of causal luck too. The explanatory luck objection is, however, more than just a hybrid of these other sorts of luck. It is its own form of moral luck as the objection argues agents are not the ultimate source of why they do what they do because what they do is just a matter of luck.

The discussion of luck that Nagel and Williams have introduced raises concerns about luck undermining control and without control our responsibility is vitiated as the control principle states. Mele is focused on luck undermining ultimacy as events caused by luck imply that agential features are not the reason for the occurrence of an event. If a choice is made because of luck, then nothing about an agent explains why an event occurs and if nothing about an agent explains why an event occurred, then they are not the ultimate source. It is possible therefore to spell out the luck objection in terms of both control and ultimacy. Luck would seem to loom large and doom the libertarian project.

³³⁰ Mele, "Kane, Luck, and the Significance of Free Will," 183.

The control principle that Nagel and Williams raised above applied to SFAs suggests that if the only difference between Joe and his action in identical worlds is luck, his action is not morally assessable because it is out of his control. Events that are lucky are beyond our control because they just occur or do not occur. Of course, we can act without luck necessarily vitiating control as some of the earlier cases like the baseball player illustrate. Luck plays a role but agents still intuitively have control in a meaningful way over what they do. In situations like the Joe case, the thought seems to be that luck plays the most important role in producing the outcome. As luck has more weight in the action Joe performs, luck undermines Joe's control.

This ties into the luck objection with regards to ultimacy. Ultimacy seems to be incompatible with luck because features of an agent are not the most significant reason as to why an agent does what they do. Nothing about Joe's character, state of mind, or desires makes the difference to the outcome. What makes the difference is luck and hence Joe is not ultimately responsible for or the ultimate source of what he does.

Consider Mele's case about Bob: in a possible world *W1*, Bob lives in a small town where they make bets on strange things like whether the opening coin toss of a football game will happen on time. After Bob agrees to flip the coin at noon, Carl, a notorious gambler, offers Bob \$50 to delay the coin toss to 12:02 pm. Bob has a genuine moral dilemma as he is resistant to taking the bribe but he is also tempted by the money. Bob takes the bribe and pretends to not know where the coin is at noon and tosses it at 12:02pm.³³¹ Mele goes on to describe the issue in the following lengthy passage:

Imagine a genuinely indeterministic number generator. At five-minute intervals, consistently with the past up to the pertinent time and the laws of nature, it can generate any one of

³³¹ Alfred Mele, "Free Will and Luck: Compatibilism versus Incompatibilism," *The Monist* 103, No.3 (July 2020): 263.

several numbers or no number at all. Its generating the number 17 at *t* is one possible continuation of things, and the same is true of several other numbers. At noon today, the machine generates the number 31. After you verify that, you might find yourself with the following belief: the machine's generating the number 31 was a possible continuation of the past up to noon, and that continuation actually happened at noon. If you were somehow to verify that, at noon, Bob decided to toss the coin at 12:02 and to pretend to be searching for it in his pockets in the meantime (decided to cheat, for short), you might find yourself with a parallel belief: Bob's deciding to cheat was a possible continuation of the past up to noon, and that continuation actually happened at noon. As I mentioned, typical libertarians contend that Bob's being directly morally responsible for deciding to cheat and his directly freely deciding to cheat require that at least one other continuation was possible at noon, a continuation in which Bob does something else at noon. Suppose that another possible continuation was Bob's deciding at noon to toss the coin straightaway; in another possible world with the same past as *W1* up to *t* and the same laws of nature, that is what happens. As I have observed more than once (Mele 2014, 547; 2017, 113–14), this supposition may be viewed by some readers as a double-edged sword. A philosopher may believe that having control over whether one *A*-s or does something else instead is required for directly freely *A*-ing and for being directly morally responsible for *A*-ing and believe that having such control requires that *A*-ing at *t* and doing something else instead at *t* are possible continuations of the past up to *t* for the agent. And the very same philosopher may worry that these possible continuations are similar enough to possible continuations for the indeterministic number generator that whatever control the agent may have over whether he *A*-s or does something else instead falls short of what is required for directly free *A*-ing and for direct moral responsibility for *A*-ing. Given the details of Bob's story, how can Bob have enough control over whether he decides to cheat or does something else instead at noon for his decision to be free and for him to be morally responsible for deciding to cheat? This is an instance of the central question posed by what I called "the problem of present luck"³³²

As we can see at the end of the passage, Mele is concerned about control. In the above passage the emphasis is on luck and control and not about features of Bob producing his *A*-ing. Mele is asking for an account of control according to which agents have control over undetermined actions. Mele's formulation of the luck objection utilizes a contrastive explanation that has some property of the agent in question that explains the difference in outcome. Consider the following operative definition of contrastive explanations: contrastive explanations are where the explanandum (fact to be explained) is why *x* rather than *y*.³³³ The General turns the key in the

³³² Mele, "Free Will and Luck: Compatibilism versus Incompatibilism," 264-265.

³³³ Fred Dretske, "Contrastive Statements." *The Philosophical Review* 81, no. 4 (1972): 420.

ignition of his car and the car explodes. A contrastive explanation explains or answers why turning the key caused the car to explode rather than starting the car.³³⁴

Mele's articulation of the challenge to control from luck is similar to the earlier discussion on Nagel and Williams. Nagel and Williams articulate all the ways our lives are touched by luck and how we might have far less control over our actions than we might pretheoretically believe. Mele is more narrowly focused on luck introduced by requiring indeterminism for agents to have control over their behavior. Mele's concern can be sharpened by noticing that the account of control he wants requires a particular kind of explanation. What is required is a contrastive explanation of undetermined actions. We need to be able to explain in contrastive fashion why Bob, given Bob's story, chooses to lie rather than the alternative. The move here is that without the contrastive explanation of an agential sort, the agent lacks control over their undetermined action. The vitiation of ultimacy is caused by the source of the event not being the agent but instead, luck.

Mele's Bob case presents the luck objection as a concern over control: "Given the details of Bob's story, how can Bob have enough control over whether he decides to cheat or does something else instead at noon for his decision to be free and for him to be morally responsible for deciding to cheat?" In the earlier passages Mele was concerned about ultimacy as he wanted a feature of an agent that would be able to account for the difference between identical agents. Rather than trying to fight both fronts simultaneously I am going to focus on ultimacy. Through focusing on ultimacy I will demonstrate that luck does not vitiate ultimacy and ultimate

³³⁴ The Dretske example is drawn from his discussion of structuring versus triggering causes and their connection with contrastive explanations. I want to make clear I have no commitments to a particular model of contrastive explanations or Dretske's distinction between triggering and structuring causes. I am only providing a broad definition.

responsibility. If this is true, then the control problem may be side-stepped as being ultimately responsible and an ultimately source is the fundamental idea that delivers libertarianism.

3. Ultimacy and Luck

I will now introduce two cases that involve luck but in which the agents are still appropriately thought of as an ultimate source of their actions. Consider the following two cases about two individuals named Alfred and Brenda:

Alfred uses a coin toss to decide between two options. The outcome of the toss is either heads or tails—this is not up to Alfred. But he endorses this method of choosing, including the outcomes that result. And he initiates the operation of this way of choosing—i.e., he chooses when to use this method, and he tosses the coin.

Brenda lives in Ottawa. She decides to live one whole year there wearing nothing but short-sleeved shirts and shorts (and sneakers on her feet). Part of her reason for doing this is to experience the effects of harsh winter weather on such a slightly dressed person. She doesn't know how cold the winter will be exactly, and she doesn't know if she'll have to go out on the worst of days. What the weather will be is not up to Brenda, nor are the effects on bare skin. But she endorses the exposure that she will undergo, and to a certain extent her engagement with the weather is up to her, since she can avoid going outdoors sometimes.

While Alfred and Brenda are engaged in activities which involve luck the outcomes of their decisions—for example Alfred's choice of hamburger for dinner using the coin toss, Brenda's discomfort on some days, but avoidance of frostbite on others, even when she goes outdoors—is attributable in some sense to Alfred and Brenda, maybe even wholly. Recall from the discussion in chapter four that attributability is a condition for ultimate sourcehood. If it is correct to attribute the results of the relevant lucky processes to Alfred and Brenda, then it follows luck is not necessarily a threat to ultimacy.

The reason for this is that Alfred and Brenda stand in the right relationship to the activity that involves luck. The nature of this relationship that allows for Alfred's action to be attributable to him is one of endorsement. Alfred endorses both using this tiebreaking method to choose and

when he uses this method to make decisions. Similarly, Brenda endorses her exposure to cold weather and obviously has no control over the weather. Yet her willingness and endorsement of exposure to cold weather makes this activity attributable to her. Alfred's circumstances involve choosing a tie-breaker when his reasons are not conclusive on their own. His endorsement of the outcome of the coin toss suffices for the action to be attributable to him as the source of it. He chooses this method, and while the coin toss itself is not within his control, his endorsement of the outcome and the coin toss seem to make him the source of the outcome.

Alfred and Brenda's endorsement places them in the right relationship to the lucky processes they are in relationship to because of their willingness to use these processes. Alfred and Brenda both act for reasons that are voluntarily, intentional and not produced by coercion. While the outcome of their activities is subject to luck, they willingly do these things and the lucky process involved produces things they each endorse. The reasons involved in the case are separate from the endorsement, but they are relevant to the description of the cases. They are important for demonstrating that Alfred and Brenda are acting freely in the sense that they are acting intentionally and on purpose; it is significant for establishing that their actions are not just an accident or purely random. The role of endorsement in their respective cases is to show how the action they perform is still attributable to them in the sense of ultimacy. They endorse the outcome of the lucky process they choose and endorse utilizing lucky processes. The endorsement of the agent is the bridge between agency and luck that allows the compatibility of ultimacy and luck.

Let's slightly alter Alfred's case to be about specific choices: whenever Alfred is faced with a torn choice with two options that rationally satisfy, Alfred tosses a coin. A choice rationally satisfies when the reasons in support of making that choice are not sufficient to rule

out an alternative as unreasonable or undesirable. Luck may be involved but the details of the case do not rule out the compatibility of attribution and ultimacy with luck. Moreover, the luck in Alfred's case is brought into play by Alfred in the sense that he chooses to use a coin toss to be a decision maker and could have chosen something else. He endorses using it and decides when to use it and it is also true that the outcome of the coin toss is lucky. While Alfred's choice involves a lucky process, his endorsement of this process makes it attributable to him. He does what he wants to do regardless of what the outcome is, and his action is not accidental or contrary to his will. Thus, Alfred stands in the right relationship to the outcome of the coin toss.

We can dig deeper into the nature of this relationship and, I think, it is a relationship we should be pluralists towards. It is possible for an agent to stand in relation to luck in at least a few different ways. Consider the following three descriptions of the possible relationships an agent can stand in with regard to luck:

1. An agent can stand in the right relationship to luck through the awareness of its presence and the explicit endorsement of it.
2. An agent can stand in the right relationship to luck through an awareness of luck involved in the production of luck. An agent need not endorse luck in this scenario and might disavow luck and fight against it.
3. While the first options involved agents having an awareness and either an endorsement or disavowal of luck, a third option it seems to me involves luck being part of the content of an agent's self-image of one's psychology and specifically action-producing processes use when they do their work. This need not arise to the level of awareness of the person whose actions are governed by such a self-image.

As I said, I do not see why we must say one of the three options is only the right approach to being a relationship to luck that does not undermine luck. Before defending this point though, it is necessary to show that separately each of the options do in fact preserve agency and specifically ultimacy. Beginning with option one, Alfred and Brenda are cases that illustrate this relationship. Alfred and Brenda's endorsement demonstrate their willingness to undergo activities—the coin toss in Alfred's case and Brenda's exposure to cold weather while being underdressed—which involve luck. They are fully aware of the luck inherent in their actions and willingly pursue their goals with their respective lucky action processes. In this relationship between agents and luck, agents can see that the outcome of their actions are uncertain and open to being influenced by luck. Nonetheless, the outcome is something they want to do and so they endorse the outcome even though luck is involved in its production. It seems to me the kind of luck in cases like Brenda is of the nature that agents can be fully aware of its presence and fully willing to accept its presence and role. Coin tosses and the weather are things whose outcomes mature agents are aware to be beyond their control but still willingly endorse engaging with or using to settle their minds. Brenda knows full well what Ottawa is typically like in the winter and she does not control the weather but still chooses to go out in summer clothes. Alfred does not control the outcome of the coin toss but endorses whatever outcome it 'decides.' He need not use a coin toss, he could have chosen some other means he might have control over. He intentionally uses a coin toss to decide for him. That these agents willingly, voluntarily and intentional engage in these lucky processes makes it appropriate to attribute what they do as ultimately up them and ultimate sources of their behavior.

The second option states an agent can be in a relationship to luck where they merely acknowledge the existence of luck in the outcome of their action. How would awareness of luck

without explicit endorsement establish the right relationship? I argue that being aware of the influence of luck is sufficient for an agent to be the ultimate source of the choice. Imagine that instead of Alfred and Brenda endorsing the use of lucky processes, they merely acknowledge that luck is involved. Suppose Alfred uses the coin toss but he does not endorse using this as a method but indifferent. Alfred is simply lazy and uses a coin toss for convenience. Alfred is aware that by using a coin toss he opens himself up to luck. Through his awareness that using a coin toss subjects him to luck and choosing to go through with it, this makes Alfred in the right relationship to luck.

Imagine Alessio making a choice we will say he is torn between. His reasons rationally satisfy making either alternative psychologically open to him. Suppose that Alessio is aware that his reasons motivate him in either direction and so he thinks, rightly, that what choice he ends up making will be open to luck. Alessio need not endorse the luck here, he might even disavow it. But his willingness to try to accomplish his goal even though his ends are subject to luck makes his choice attributable to him. Much like how an agent can still be responsible for an action that is undetermined because they still contribute to the outcome, Alessio contributes to the outcome through his choice and his acknowledgment of luck makes his choice attributable to him.

Now let's imagine a way to be in the right relationship that does not require being aware of luck. This has the advantage of being a category that is wider than the previous two because they did require an awareness about luck and either an endorsement or acknowledgement. An agent can be in the right relationship to luck here without an awareness or concern about luck. All that would be needed is a view of oneself and one's actions as subject to luck to be incorporated (perhaps in specific ways) into one's sub-personal self-image-producing and action-producing

capacities. In this case, we have the possibility of all normal people acquiring such a self-image, and hence the right relationship to luck in general, through typical maturation. The (by hypothesis uncommon) absence of such a sub-personal self-image might be detectable by things like strange sorts and/or degrees of surprise about worldly processes that affect one's actions in unusual ways, or lack of flexibility when the world at large doesn't meet one's expectations. Such responses (behavioral, attitudinal) would mark a dissociation between one's view of oneself and one's understanding of the wider world and the contingency found therein.

The advantage of this view over the previous two is above mentioned possibility of all normal people developing the right relationship to luck. This seems like a virtue of an account that seeks to make ultimacy compatible with luck. This also opens the possibility that any normal people that don't develop this sort of relationship to luck could end up in the relationships described in the first two cases. Indeed, if one saw one as subject to luck but disavowed it in the maturation of oneself, this would put an agent in the second sort of relationship. If one came to be aware and endorsed luck, an agent would be in the first sort of relationship.

Suppose Amara is a maturing agent who is making a choice between rationally satisficing reasons. This introduces luck into her deliberative and action-producing processes. For Amara to stand in the right sort of relationship to luck she need not be aware, or care about luck. What is needed is a view of herself as being open to not luck per se as this would bring luck to a level of awareness not required. Instead, Amara needs to come to incorporate luck into her sub-personal self-image through the development of a self that sees herself as open to failure and an uncertainty about what to do and a willingness to choose, nonetheless. What I am trying to capture here is that Amara sees her choice as being open to not meeting her expectations and that she doesn't know which choice is the right one to make. The choice she ends up making might

be subject to luck, but from the first-personal perspective she doesn't see herself as making a lucky choice. She sees herself as making a decision she is not certain is the right one but which she willingly and intentionally chooses. At the sub-personal level her action-production might involve luck, but this does not vitiate the attribution of her choice to Amara. Amara's choice still comes from her in that she is choosing for reasons that motivate her and this would be true no matter what Amara does. That luck is involved seems besides the point to Amara developing into a mature agent which, in the view of my theory, is an ultimately responsible self-author. Amara at the level of her awareness is making choices that are importantly connected to her will: her values, desires and reasons are all causal contributors to production of her choice. In spite of the luck at work here, Amara is still an ultimate source in virtue of the contributions of herself in the form of these agential features make. The outcome of this process would have been subject to luck but as long as Amara is involved in this manner, she has the necessary relationship to luck to say she is an ultimate source.

4. The First SFA and Luck

Rosa is making her first SFA when she attempts to take a cookie when her parents tell her no more cookies. Imagine a cross-world agent, Rosa*. Rosa and Rosa* are identical, and in this case they have no prior history or fully structured narrative to guide their choice. In this case, how is the choice or action attributable to Rosa as the ultimate source? Rosa is not born the author of her life story as it literally begins when she is born but she becomes an author through her SFA. When Rosa comes to her first SFA, this marks a change in the relationship between herself and her life through performing a novel decision, choice, action etc. The performance of an SFA for the first time marks that the way Rosa relates to her life is beginning to transform in new and important ways. Understanding her emerging relationship with SFAs and life narrative

will be how to show that Rosa makes an SFA without luck undermining the attribution of ultimacy to her.

The first step in describing this relationship is to bring in certain aspects from chapter five regarding the formation of our narratives. I discussed how Nelson described the way in which children start out as non-authors or narrators and become authors. It is relevant to recall that Nelson hypothesized “a new level of consciousness that emerges in early childhood together with a new sense of self situated in time and in multiple social realities.” I hypothesize that as part of this new level of consciousness and sense of self, Rosa feels a new sensitivity to reasons for action where she can feel torn over what to do. Consequently, Rosa exercises a new kind of agency that comes from this new relationship to herself, and her life. She does not have much of a past, but she is aware she comes from somewhere as she knows she has parents. Her sense of the future is tied to her present decision to either listen to her parents or take a cookie contrary to her parents wishes. While Rosa does not have a view of her choice extending her past into the future, her temporal narrative self emerges in the choice made in the first SFA. As every SFA is a value experiment and this is her first SFA, hence her first value experiment, the relationship she has to her first value experiment is illuminating for attributing it to her as the ultimate source of it.

Value experiments per se do not mitigate the issue but it is what making them expresses and how beginning to initiate them marks the beginning of the change of oneself and our relation to our lives. Making a value experiment expresses a relationship to one’s life where one is willing to take responsibility for initiating a novel process or to go down a particular forking pathway. Nelson’s view of the development of children is significant for present purposes as it gets at the relationship one has to one’s life and how it changes in important ways. When Rosa is

able to make her first SFA and hence first value experiment, this denotes the beginning of the development of the new sense of consciousness and sense of self. When she is able to make such a choice for the first time she is able to recognize reasons. She sees that she has reasons to do A and different reasons to not A, and instead perform B. She can see she wants to both but cannot and endorses making a choice to do either A or B for those reasons. She appears to be able to act as an agent and starts to exercise authorship over her life. Up until now, Rosa has been living in a story she has been a mere character of, but now that she is able to make a choice, she endorses for reasons that, she is starting to be the author of the story and character that is herself.

The relationship between Rosa and her first SFA that makes the outcome ultimately hers is Rosa's new attitude toward her choice. As she is still a young inexperienced agent her sensitivity is crude and so her responsibility is quite limited. Her awareness of herself makes this choice hers as she can see she can make a decision in ways she could not before. In her first SFA, she desires two conflicting ends and has to try to resist the temptation posed by the alternative. She is actually thinking about her reasons in a way she did not before: as a young immature child she just did what she wanted. Now, she actually feels a hesitation from her emerging sensitivity to reasons between different kinds of reasons. She can for the first time be an author as she is able to be sensitive enough to reasons and make genuine choices in a way that separates Rosa from younger children that have not reached this level of development. Rosa's attitude towards the choice is characterised by a willingness to do either for reasons that satisfy. When Rosa was younger, she would react to her environment in a way that was not sensitive to reasons in the same way as her dilemma. Before now, when one event happened Rosa would respond such that we would not want to say she is sensitive to the reasons to respond to. Suppose Rosa would not think twice about being cut off from extra cookies, but this time she is actually

sensitive enough in responsiveness to reasons that she actually contemplates disobeying her mother. The hesitation in her response to the commandment is a sign of an emerging self-consciousness as she can consider an alternative to the option presented to her. She considers stealing the cookie contrary to being told not and this is the first time this thought occurs to Rosa. Prior to this, she would abstain merely because she would listen to her mom.

Rosa I think captures the third kind of relationship one can have to luck I briefly introduced in the previous section. Rosa doesn't think about luck in her deliberations over stealing cookies but agents need not be aware of luck for them to be in the right relationship. One only has to incorporate luck into one's sub-personal self-image-producing and action-producing capacities. Rosa, I think, achieves this by her willingness to act and endorsement of her action even though she knows she might get caught or be unhappy. Since an agent does not know what will happen in an SFA there is a kind of luck involved here. Being willing to choose with this knowledge incorporates luck in the appropriate way for Rosa to meet the condition and have the product of her first SFA be attributed to her.

5. Self-Formation and Luck

Consider Mele's case about Ann and Ann*: Ann has made a promise to call her colleague Beth at eight in the morning to prepare her for an important meeting. Beth has recently broken a promise to Ann and Ann is still upset over it. Ann believes she ought to call Beth on time but she also wants revenge. She is conflicted and engages in an SFA where she chooses to get revenge and does not call Beth on time and waits several minutes. Consider an agent in an identical possible world and who is identical to Ann—call her Ann*—who faces the same situation. She too faces the same internal conflict but is able to overcome her desire for revenge and calls Beth

on time.³³⁵ Mele challenges the libertarian to explain this difference in outcome. Mele wants an explanation that cites the, ““powers, capacities, states of mind, moral character, and the like that explains the difference in outcome” Without an explanation that uses something along these lines “then the difference really is just a matter of luck.”³³⁶

I will grant that the difference in outcome is luck. The Anns are an articulation of the luck objection focused on ultimacy. Given that Ann and Ann* can do otherwise voluntarily, intentionally, and purpose what is it about them that explains why they decide differently from each other? Since they are identical there is by hypothesis no such reason and hence the reason for the difference is luck. The proponents of the luck objection think the conclusion to draw from the case is that agents are not an ultimate source or ultimately responsible. We could also draw the conclusion that agents have no control because of the absence of the right sort of explanation. In addressing concerns over luck, I will address ultimacy and control separately and begin with ultimacy.

The Alfred case showed that if an agent has the right relationship to an activity involving luck, ultimacy is preserved. I am going to apply that lesson to the Anns case. The way this will be done is by bringing my narrative account into the relationship between the Anns and their SFAs. Ann and Ann* have a life story and this is important because it shapes their current situations. Ann’s narrative structure has developed in such a way that it leads to the two options before her, and her life story does not present a sufficient reason to only pick one option; this is true of Ann* too. The Ann’s life stories generate satisficing reasons and her reasons make it possible for them to do otherwise. The Anns share with Alfred that their action involves luck but

³³⁵ Mele, “Kane, Luck, and the Significance of Free Will,” 98.

³³⁶ Mele, “Kane, Luck, and the Significance of Free Will,” 99.

I am going to argue this does not undermine ultimacy. The narrative relationship the Anns have with their SFAs plays an important role in demonstrating this alongside their endorsement of the outcome of these situations. Endorsement played a big role in the Alfred case in showing how ultimacy is consistent with luck and in the Anns case this is true too. Endorsement, however, in the Anns case is situated within their life stories. There are specific features of being the author of one's life story that are related to endorsing what one does when making an SFA.

One significant feature is that when an agent makes a choice, they are making a value experiment. When agents make a value experiment in the context of an SFA, their SFA is in a wider context of their life story. To make a value experiment is to choose to extend the life story in one way rather than another. As the options rationally satisfy, the Anns are phenomenologically and causally equally attracted and repelled by their options. So, they will endorse the outcome of their SFA as a consistent and intelligible continuation of their life story. Since their choice is experimental, they are taking a risk in an ordinary sense of not knowing the outcome in advance, as well as taking a risk in pursuing an option that might go contrary to their expectations. Both the choice itself and the options are importantly linked to the narrative structure of an agent. This constrains the outcomes of an SFA though it does not make the difference to what occurs. It is important nevertheless to keep this in mind as it shows that agents are not open to just any possibility. The constraint of the situation placed by the narrative structure of their lives is important because the very dilemma that seems to open the door to luck is generated by an agent's values, beliefs and reasons. The structure of one's will is not set so one finds oneself in a predicament between attractive alternative possibilities. This is important for showing how luck is consistent in the Anns case with ultimacy. The luck involved in the Anns situation is as a result of their finitude. The luck is further constrained the Anns narratives:

it's luck whether option A or option B is chosen, but these are the options because of the life story of the Anns to this point. As agents in an SFA will see the outcome of SFAs as a meaningful continuation of their narratives either way the outcome is endorsed by them and hence attributable to them.

In the Anns case, the endorsement that delivers the compatibility of luck and ultimacy derives from the fact that the Anns see either option as intelligible extensions of their life story. Since either option are related to their pasts and intelligently extend their presents into their futures, the Anns endorse the outcome. Luck does not prevent their SFAs as being attributed to them because Ann and Ann* endorse them as the meaningful continuation of their life story. Given their psychology, the Anns' pasts lead to certain possible futures and whatever they do they could have done otherwise. While the difference between Ann and Ann* is luck their SFAs are connected and attributed to them because of phenomenological and causal connection to each of their narrative structures.

A second feature of my view of narrative self-formation relevant to the present topic is the nature of the reasons and causes involved. Recall on my view of narrative that there are necessary, sufficient and satisficing reasons/causes. Ann and Ann* have their narratives generate satisficing reasons which is why they are able to do otherwise rationally and intentionally. Ann and Ann* engage in value experimentation precisely because of the fact that their reasons satisfice and are not sufficient. They are also making an SFA and all SFAs are value experiments. Value experiments have uncertain outcomes and so there is another kind of luck different from the sort concerning Mele. There is the luck involved in which value experiment is chosen and then there is the luck involved in what happens in a value experiment. The concern about value experiments and luck is that, since agents do not know what will happen, the

outcome is just a matter of luck and so not within their control. The luck Mele is concerned about is which value experiment will be run, as without an agential explanation of why one value experiment rather than another, the choice is just a matter of luck and hence not within one's control. Since what happens in a value experiment is just a matter of luck, agents do not control what happens in SFAs, and since these are the choices where we shape our selves we don't control them and hence are not responsible for them. I do think the two threats from luck are related and a solution to one can be used to mitigate the other.

The luck that seems to be involved in value experiments found in SFAs surrounds the outcome not being guaranteed to happen because it is undetermined. It is also uncertain whether from the agent's point of view they will regret their decision or feel like it was the right thing to do for them. We can marshal the reflections on control and reasons to mitigate this issue. Value experiments have an uncertainty to them, but this is true whether or not determinism is true. Value experiments include both determined and undetermined actions and involve uncertainty in either case. However, in cases of SFAs, the outcomes are not unlimited or purely random. The businesswoman can help or go to her meeting, and it makes sense for her to do either, but it would be completely random for her to hop on a plane to Italy. Her narrative past limits the outcome of her experiment to either trying to help the victim or trying to get to her meeting. If any other event occurred, this would be just an accident or something the businesswoman is not free to do by any of the standards I have laid out.

The luck regarding the outcome of a value experiment seems to me to be less of a challenge than the choice of value experiment. First of all, it is the agent's narrative structure and character that generates the viable alternatives. For instance, the businesswoman is going to help the assault victim or go to her meeting, she is not going to get on a plane to Italy. The outcome of

any particular value experiment while uncertain is still causally constrained by the inputs. Since value experiments don't have to be restricted to SFAs, there are value experiments which are determined but which still have this uncertainty, and so this line of thought would suggest that even if determinism were true value experiments are just a matter of luck. This would be a problem of luck for compatibilism because the luck here is generated from uncertainty; there is no mention here of something like ultimacy which compatibilism does not care about. Perhaps the thought is that if an action is undetermined then the outcome is uncertain in a metaphysical sense for a determined value experiment it is only epistemic. It is true that when making an SFA we are trying to choose different value experiments, but we must keep the uncertainty over which experiment is chosen separate from the uncertainty of the agent's phenomenology. The former is where the luck introduced by Mele needs to be addressed while the latter is a pretty standard feature of everyday life. Citing uncertainty as a reason for the outcome being pure luck seems like a red herring. Putting aside the concern about luck for the choice of value experiment, value experiments when chosen are not open to being just a matter of luck in the sense of the outcome merely occurring independently of the agent's psychology. The businesswoman can only freely and responsibly choose between the alternatives in the circumstance that are supported by her satisficing reasons and other agential aspects like her values and the kind of person she is. Her agential features do not guarantee an outcome, but this seems to me not to be an issue. The assassin could not guarantee that he would hit his target but it does not follow that because he might have failed he is not the source of the event of the prime minister dying.

It does not follow that because one cannot guarantee an event to occur before it occurs that one does not control or produce an event when it occurs especially when one makes efforts to make the event occur. The fact remains that it is one's will which generates what possibilities are

motivating to an agent; whatever motivation is pursued one remains the ultimate source of the outcome of the event. If we are the ultimate source of the outcome of a value experiment, it seems possible to build a bridge to control but we need to build the connection between ultimacy and being the source of undetermined events.

Value experiments like SFAs seem to have these features where the relationship between the outcome and the agent are such that the indeterministic process does not rule out ultimacy or ultimate responsibility. The employee and the karate master through their efforts and intentions are trying to bring about an event. and if this occurs the mere fact that it might not does not undermine their control. The reason this is true is that the events would not occur without the agents trying to produce them. Of course, it is possible for the event to still occur but at the time of the event, without the employee's evil intentions and the motion of the master's arm swing, the outcome would not occur when it occurred.

The discussion here points to the importance of an agent in the production of an event even when it is undetermined. All of the cases share the characteristic of an agent trying to do something and endorsing the outcome of their efforts. In cases of SFAs, agents are trying to make up their mind between attractive alternatives coming from their narratives. The businesswoman and the Anns are engaged in a value experiment where the outcome continues their life story one way rather than another.

When an agent chooses their value experiment, they can be thought of as essentially saying:

True enough. But I did have adequate reasons for choosing as I did, which I'm willing to endorse and take responsibility for. If they were not sufficient or conclusive reasons, that's because, like the heroine of the novel, I was not a fully formed person before I

chose (and still am not, for that matter) ... I am in the process of writing an unfinished story and forming an unfinished character who, in my case, is myself.³³⁷

The reflections on luck and ultimacy mitigate the force of the luck objection. It is not the case that ultimacy is inconsistent with luck in the absence of control. Agents as self-forming authors use value experiments to decide how to develop their narratives, and while the outcomes of SFAs are lucky, agents are still ultimate sources of the outcome of their value experiments. Since either value experiment fits with their overall narrative, they willingly endorse the outcome.

I have argued that SFAs are attributable to agents as ultimate sources because of the relationship between them and the outcome of SFAs. I argued that agents have a narrative structure that ties the outcome of SFAs to agents in virtue of agents endorsing their outcomes as intelligibly continuing their life story. This captures an aspect of the kind of libertarianism I am interested in defending as being what Kane describes as a “teleological intelligibility” libertarian theory. As a helpful reminder, teleological intelligibility means that SFAs can fit into “meaningful sequences... or fit into a coherent narrative for which the agents themselves are at least partly responsible and in which they take responsibility for the novel pathways they initiate.”³³⁸ I have described in the discussion so far how agents take responsibility for the novel pathways and their narratives deliver the right relationship between agents, their SFAs and luck.

In summary, I have demonstrated that Mele’s concern about luck can be mitigated and does not undermine libertarian free will. I have outlined how even though SFAs are lucky processes the ultimacy necessary for free will and moral responsibility is not vitiated. Through contextualizing an agent’s SFA as making a value experiment and placing an agent’s SFA/value

³³⁷ Kane, “Acting of One’s Own Free Will,” 53.

³³⁸ Kane, “Acting of One’s Own Free Will,” 53.

experiment in a narrative structure, an agent retains ultimate sourcehood and ultimate responsibility for the outcome of the lucky process. Thus, I have shown how to undermine the force of the problem of luck that Mele introduced at the beginning of the chapter.

Chapter 7

Conclusion

1. How We Got Here

I stated in chapter one that my view of free will focuses on the power of agents to be the authors of one's own life story. I explained in chapters four and five that we exercise this power directly and indirectly. We directly author our life stories in moments where the future remains open to us and we can shape the structure of our narratives. I said that these moments are Kanean SFAs. We indirectly author our life stories when we act from our freely shaped wills and this implies that in these particular instances we can be responsible even if we could not do otherwise. In the same two chapters I also talked about ultimate responsibility and authorship, respectively. I described how we are ultimate sources and what this means for authorship. We have the power to be the authors of our lives because we are the ultimate source of them: the general shape and structure of our life and who we are, while shaped in many ways, is significantly formed by our SFAs. At the very least, I argued, our free choices have to influence and shape who we are if we are ultimately responsible. I discussed in chapter two that Kane is correct to frame the free will debate in terms of who and what we are. I defended Kane's theory from other libertarians. I attempted to respond to the most pressing issue libertarian theories face, generally speaking; the problem of luck. I addressed Mele's formulation of the objection and claimed that, contrary to his attack, luck does not undermine ultimacy.

2. Where We Are

One of the most interesting philosophical trends to have emerged in the twenty-first century is the subfield of experimental philosophy. Experimental philosophy re-examines age-old

philosophical topics from a fresh new angle. Taking the data collection methods of the sciences and polling the views of ordinary people (non-academics) and experts, experimental philosophers try to find out what people think regarding a specific topic. There has naturally been work done related to free will on what ordinary people think free will is and its connection to moral responsibility.

The most influential work on this has been done by Joshua Knobe and Shaun Nichols.³³⁹ Their research has shown that people think of free will in both compatibilist and incompatibilist terms. What is interesting about their research is that it tracks with the concession I made earlier regarding the ability to do otherwise. I do not believe we can never be responsible unless we are able to do otherwise. The reason I gave for this is that we are responsible even when we cannot do otherwise in so far as we are responsible for the kind of person that we are. It is the actions whereby we become who we are that require alternative possibilities. Recall from chapter four the distinction between responsibility for expressing your will and responsibility for having the will you have. The incompatibility of ultimate responsibility and determinism has to do with the second sort of responsibility while the first kind is perfectly compatible with determinism.

There is data on ordinary folk intuitions that gels with this line of thinking. Recent work done by A. Shariff into the retributive attitudes of judges and jurors has shown that when people are educated about the free will debate and the neural basis of behavior, their beliefs change. They come to favour a less retributive approach as they come to believe criminals are less responsible for what they do than they originally thought. He writes,

That mere exposure to modern neuroscience can be sufficient to reduce retributivist motivations may be particularly relevant to court cases. The explicit existence of free will

³³⁹ Shaun Nichols and Joshua Knobe 2007, "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions," *Noûs* 41, no. 4 (2007): 663–85.

may be rarely debated in court, but neuroscientific evidence often is. Indeed, recent research showed that judges afforded shorter sentences to hypothetical psychopathic criminals when the description of the criminals' psychopathy included a biomechanical component, compared with when it did not (Aspinwall, et al. 2012). Our findings likewise suggest that merely presenting such a perspective may move judges and jurors toward being less punitive and less retributive in general.”³⁴⁰ (Shariff, et al. 2014, 1569)

This affirms that ‘free will’ is not a simple term with a univocal meaning. People think of freedom not as an all or nothing concept but one that can come in degrees. The evidence in the cited passage shows that people are sensitive to the influences of people’s behavior, and that if the behavior is generated by causes beyond an agent’s control they become less punitive in their reactive attitudes. This lends support to the incompatibilist’s case in that it shows that ordinary persons’ responsibility take into account how much of a role an agent played in becoming the sort of person they become. I conjecture that this shows people separate responsibility for expressing one’s will and responsibility for having the will that is expressed in action. As I have argued throughout the thesis, to be responsible for having the will one has requires having contributed to making it be the way it is. The folk intuitions presented show that people think responsibility for having the will one has is tied to what one did and did not do. In terms of my own view, I understand the data to support the importance people place on being the authors of their life stories. If who and what we are is shaped exclusively by deterministic influences and our choices play no role at all in becoming who and what we are, we do not author ourselves and our stories. This intuition is shared by ordinary folks though the evidence does not frame this in terms of authorship. Nonetheless, ordinary people appear to believe that if what we do is ultimately the product of factors beyond our control we are less responsible.

³⁴⁰ Shariff, A. F., Greene, J. D., Karremans, J. C., Luguri, J. B., Clark, C. J., Schooler, J. W., Baumeister, R. F., & Vohs, K. D, “Free Will and Punishment”, *Psychological Science*, (2014): 25.

One of the most important aspects to the free will problem, from my perspective is, the issue of meaning. By “meaning,” I mean the meaning of one’s life or the purpose of one’s life. Free will seems to me to be one of the most important tools human beings have in the creation of the meaning or purpose of their lives. This is an implication of my view because every undetermined free choice an agent makes when authoring their life story is plausibly seen as trying to make their life stories to make them intelligible to them. By “intelligible,” I mean appear coherent and provide an understanding of who they are and how they became who they are. Free will is a practical necessity used to live a life and hopefully one seen as worth living. In this way, I think free will as I understand it is closely tied to an image of human beings as, to borrow Schechtman’s phrase, “meaning seekers”.³⁴¹ My idea of authorship ties into this picture because the narrative structure of our lives is also tied to meaning. Recall Harold from chapter five, Harold uses his ability as an author to choose to shape his narrative into a story that serves him in providing him a life that is more fulfilling. As I see it, as the ultimate source and authors of who we are, the meaning of our lives, and the value we place on our lives would be affected by accepting the truth of determinism. For if we are not producers and authors, the picture of human beings as meaning seekers would need to be revised; if determinism were true, the ultimate meaning of our lives would be altered as we would have to abandon self-conceptions of authoring our lives in my intended sense. We would be losing something of value here but again I think this judgement comes down to a difference in truth about fundamental values.

3. Where We Are Going

³⁴¹ Marya Schechtman, “Self-Expression and Self-Control.” *Ratio*, 17, no. 4 (2004): 409–27.

In closing, the free will debate is far from over. I have argued for a view of free will where agents are the ultimate authors of their own life stories and identities. I think, however, that my view is itself a kind of experiment of the sort that I have talked about throughout the thesis. For my view has tried to show the plausibility, significance, and desirability of the kind of free will I have articulated and defended. However, this is just a conceptual analysis at this point, as the empirical reality of this kind of free will remains to be seen. I am making a factual experiment where I am taking a risk or prediction on what the future empirical science will say. I bet on my kind of libertarianism being vindicated. As I have confined my theorizing to the bounds or limits of nature my view is vulnerable to being disproven by developments in our understanding of nature. For instance, if the empirical thesis of determinism is proven true my theory of free will as stated is ruled out. Despite having argued for a theory of incompatibilist libertarian free will, I believe the view should be revised instead of abandoned or the contrary free will skeptic view endorsed. Free will is an integral part the descriptive image of human beings and, I think, it is a mistake to abandon it. As Sartre famously put it, human beings are “condemned to be free.”³⁴² Whatever the winds of fortune bring, I think this is an apt description of human beings. The issue it seems to me is the nature of this freedom and my thesis has tried to contribute to its understanding.

³⁴² Jean-Paul Sartre, *Being and Nothingness*, transl and eds. Hazel Estella Barnes, (New York: Washington Square Press, 1992).

Bibliography

Austin, John Langshaw. "Ifs and Can's." In *Philosophical Papers*, edited by James Opie Urmson and Geoffrey James Warnock, 153-180. Oxford: Clarendon Press, 1961.

Anscombe, Elizabeth. "Causality and Determinism." In *Agency and Responsibility: Essays on the Metaphysics of Freedom*, edited by Laura Wadell Ekstrom, 57-76. Boulder: Westview Press, 2001.

Baier, Annette. "The Search for Basic Actions." *American Philosophical Quarterly* 8, no. 2 (April 1971): 161-170.

Balaguer, Mark. *Free Will as an Open Scientific Problem*. Cambridge, Massachusetts: MIT Press, 2010.

———. *Free Will*. Cambridge, Massachusetts: MIT Press, 2014.

Berlin, Isaiah, *Liberty*. Edited by Henry Hardy. Oxford: Oxford University Press, 2013.

Broad, Charlie Dunbar. "Determinism, Indeterminism, and Libertarianism." In *Ethics and the History of Philosophy: Selected Essays*. London: Routledge & K. Paul, 1952.

Bronte, Charlotte. *Jane Eyre*. In *The Selected Works of the Bronte Sisters*. Hertfordshire: Wordsworth Editions, 2005.

Callard, Agnes. *Aspiration: The Agency of Becoming*. New York: Oxford University Press, 2018.

Capes, Justin and Randolph Clarke. Incompatibilist (Nondeterministic) Theories of Free Will", *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/entries/incompatibilism-theories/>.

Carroll, "On the Narrative Connection," in *Beyond Aesthetics: philosophical essays*. New York: Cambridge University Press, 2001.

Carroll, Noël. "On the Narrative Connection." In *Beyond Aesthetics: Philosophical Essays*. First Edition, 118-133. Cambridge: Cambridge University Press, 2001.

Chisholm, Roderick M. "Freedom and Action." In *Freedom and Determinism*, edited by Keith Lehrer, 11–44. New York: Random House, 1966.

———. "Human Freedom and the Self." In *Free Will*, edited by Robert Kane, 47-58. Malden, Massachusetts: Blackwell Publishers, 2002.

Clarke, Randolph. *Libertarian Accounts of Free Will*. Oxford: Oxford University Press, 2003.

———. "Agent Causation and Event Causation in the Production of Free Action." *Philosophical Topics*. 1996, 24, No. 2, 19-48.

Clarke, Randolph and Justin Capes, "Incompatibilist (Nondeterministic) Theories of Free Will", *The Stanford Encyclopedia of Philosophy* (Fall 2020 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/entries/incompatibilism-theories/>.

Coates, Justin D. and Michael McKenna. "Compatibilism", *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/Archives/win2022/entries/compatibilism/index.html>.

Dennett, Daniel C. *Elbow Room: The Varieties of Free Will Worth Wanting*. Cambridge, Massachusetts: MIT Press, 2015.

———. "The Self as a Center of Narrative Gravity." In *Self and Consciousness: Multiple Perspectives*, edited by Frank S. Kessel, Pamela M. Cole, and Dale L. Johnson, 103–15. Hillsdale, New Jersey: Lawrence Erlbaum Associates, 1992.

Donagan, Alan. *Choice: The Essential Element in Human Action*. London: Routledge and Kegan Paul, 1987.

Doris, John M. *Talking to Our Selves: Reflection, Ignorance, and Agency*. Oxford: Oxford University Press, 2015.

Double, Richard. *The Non-Reality of Free Will*. New York: Oxford University Press, 1991.

Dretske, Fred I. "Contrastive Statements." *The Philosophical Review* 81, no. 4 (1972): 411–37. <https://doi.org/10.2307/2183886>.

Dworkin, Gerald. "Elbow Room . Daniel C. Dennett." *Ethics* 96, no. 2 (1986): 423–25. <https://doi.org/10.1086/292763>.

Ekstrom, Laura Waddell. *Agency and Responsibility: Essays on the Metaphysics of Freedom*. Boulder: Westview Press, 2001.

———. *Free Will: A Philosophical Study*. Boulder, Colorado: Westview Press, 2000.

———. "Indeterminist Free Action." In *Agency and Responsibility: Essays on the Metaphysics of Freedom*, edited by Laura W. Ekstrom, 138-157. Boulder: Westview Press, 2001.

———. "Introduction," in *Agency and Responsibility: Essays on the Metaphysics of Freedom*, edited by Laura W. Ekstrom, 2-14. Boulder: Westview Press, 2001

———. "Toward a Plausible Event-Causal Indeterminist Account of Free Will." *Synthese* 196, no. 1 (2019): 127–44. <https://doi.org/10.1007/s11229-016-1143-8>.

Elzein, Nadine. "The Demand for Contrastive Explanations." *Philosophical Studies* 176, no. 5 (2019): 1325–39. <https://doi.org/10.1007/s11098-018-1065-z>.

Fiorello, Alessandro. "Free Will, Values, and Narrative Selfhood." *Philosophia* 48, no. 1 (2020): 115–32. <https://doi.org/10.1007/s11406-019-00065-9>.

Fischer, John Martin. *My Way: Essays on Moral Responsibility*. New York: Oxford University Press, 2006.

———. "The Cards That Are Dealt You." *The Journal of Ethics* 10, no. 1/2 (2006): 107–29. <https://doi.org/10.1007/s10892-005-4594-6>.

———. *The Metaphysics of Free Will: An Essay on Control*. Oxford: Blackwell, 1995.

———. "Compatibilism." In *Four Views on Free Will*, 44–84. First Edition. Malden, Massachusetts: Blackwell Publishing, 2007.

Fischer, John Martin. "Sourcehood: Playing The Cards That Are Dealt You." In *Deep Control Essays on Free Will and Value*. New York: Oxford University Press, 2012.

Fischer, John Martin, and Neal A Tognazzini. "The Truth about Tracing." *Noûs* 43, no. 3 (2009): 531–56. <https://doi.org/10.1111/j.1468-0068.2009.00717.x>.

Fischer, John Martin, and Mark Ravizza. *Responsibility and Control : A Theory of Moral Responsibility*. Cambridge ; Cambridge University Press, 1998.

Fischer, John Martin, Robert Kane, Derk Pereboom, and Manuel Vargas. *Four Views on Free Will*. First Edition. Malden, Massachusetts: Blackwell Publishing, 2007.

Forster, Marc. *Stranger than Fiction*. Culver City, California: Sony Pictures Home Entertainment, 2007.

Feinberg, Joel. *The Moral Limits of the Criminal Law Volume 3, Harm to Self*. New York: Oxford University Press, 1984.

Forrest, Peter. "Backward Causation in Defence of Free Will." *Mind* 94, no. 374 (1985): 210–17.

Frankfurt, Harry G. "Freedom of the Will and the Concept of a Person." In *The Importance of What We Care About: Philosophical Essays*, 11–25. Cambridge: Cambridge University Press, 1988.

———. "Reply to John Martin Fischer." In *Contours of Agency: Essays on themes from Harry Frankfurt*. Edited by Sarah Buss and Lee Overton. Cambridge, Massachusetts: MIT Press, 27–31.

———. "The Faintest Passion." In *Necessity, Volition, and Love*. Cambridge: Cambridge University Press, 1999.

Franklin, Christopher Evan. *A Minimal Libertarianism: Free Will and the Promise of Reduction*. New York: Oxford University Press, 2018.

- . “How Should Libertarians Conceive of the Location and Role of Indeterminism?” *Philosophical Explorations* 16, no. 1 (Nov 2013): 44–58. <https://doi.org/10.1080/13869795.2013.723036>.
- Ginet, Carl. “Can an indeterministic cause leave a choice up to the agent?” In *Libertarian Free Will: Contemporary Debates*, edited by David Palmer, 15–26. New York: Oxford University Press, 2015.
- . “Freedom, Responsibility, and Agency” *The Journal of Ethics* 1 no. 1, (1997): 85–98.
- . *On Action*. Cambridge: Cambridge University Press, 1990.
- Griffith, Meghan. “Why agent-caused actions are not lucky.” *American Philosophical Quarterly (Oxford)* 47, no. 1 (2010): 43–56.
- Grunbaum, Aldoph. “Free Will and Laws of Human Behavior.” *American Philosophical Quarterly* 8, no. 4 (1971): 299–317.
- Goetz, Stewart. *Freedom, Teleology, and Evil*. London: Continuum, 2008.
- . “A Noncausal Theory of Agency.” *Philosophy and Phenomenological Research* 49, no. 2 (1988): 303–16. <https://doi.org/10.2307/2107978>.
- . “Libertarian Choice.” *Faith and Philosophy: Journal of the Society of Christian Philosophers* 14, no. 2, (1997): 195–211.
- Harris, Sam. *Free Will*. New York: Free Press, 2012.
- Hobbes, Thomas. *Leviathan*. Edited by Crawford Brough. London: Penguin, 1985.
- Hume, David. *A Treatise of Human Nature*. Edited by Ernest Campbell Mossner. Harmondsworth: Penguin, 1985.
- Hobart, R. E. “Free Will as Involving Determination and Inconceivable Without It.” *Mind* 43, no. 169 (1934): 1–27.
- James, William. “The Will to Believe.” In *The Will to Believe and Other Essays in Popular Philosophy*, 1–31. New York: Longmans, Green, and Co., 1897.
- Kane, Robert. “Acting ‘of One’s Own Free Will’: Modern Reflections on an Ancient Philosophical Problem.” *Proceedings of the Aristotelian Society* 114 (2014): 35–55.
- . *A Contemporary Introduction to Free Will*. New York: Oxford University Press, 2005.
- . “Dimensions of Responsibility: Freedom of Action and Freedom of Will.” *Social Philosophy & Policy*, no. 1 (2019): 114–31. <https://doi.org/10.1017/S0265052519000232>.

———. *Ethics and the Quest for Wisdom*. Cambridge: Cambridge University Press, 2010.

———. “Free Will and the Dialectic of Selfhood: Can One Make Sense of a Traditional Free Will Requiring Ultimate Responsibility?” *Ideas y Valores* 58, no. 141 (2009): 25–43.

———. “Free Will: New Directions for An Ancient Problem.” In *Free Will*, edited by Robert Kane, 222-248. Malden, Massachusetts: Blackwell Publishers, 2002.

———. “Libertarianism.” In *Four Views on Free Will*. 5-43. First Edition. Malden, Massachusetts: Blackwell Publishing, 2007.

———. “Moral Responsibility, Reactive Attitudes and Freedom of Will.” *The Journal of Ethics* 20, no. 1/3 (2016): 229–46. <https://doi.org/10.1007/s10892-016-9234-9>.

———. “New Arguments in Debates on Libertarian Free Will: Responses to Contributors.” In *Libertarian Free Will: Contemporary Debates*, edited by David Palmer, 179-214. New York: Oxford University Press, 2015.

———. “On the Role of Indeterminism in Libertarian Free Will.” *Philosophical Explorations* 19, no. 1 (2016): 2–16. <https://doi.org/10.1080/13869795.2016.1085594>.

———. “Rethinking Free Will: New Perspectives on an Ancient Problem.” In *The Oxford Handbook of Free Will*, edited by Robert Kane, 381-404. Second Edition. New York: Oxford University Press, 2011.

———. “The Complex Tapestry of Free Will: Striving Will, Indeterminism and Volitional Streams.” *Synthese* 196, no. 1 (2019): 145–60. <https://doi.org/10.1007/s11229-016-1046-8>.

———. “The Dual Regress of Free Will and the Role of Alternative Possibilities.” *Noûs* 34, no. s14 (2000): 57–79. <https://doi.org/10.1111/0029-4624.34.s14.4>.

———. “The Problem of Free Will: A Libertarian Perspective.” In *Do We Have Free Will?: A Debate*, edited by Robert Kane and Carolina Sartorio, 3-67. New York: Routledge, 2022.

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

———. “Two Kinds of Incompatibilism.” *Philosophy and Phenomenological Research* 50, no. 2 (1989): 219–54. <https://doi.org/10.2307/2107958>.

Kant, Immanuel, Paul Guyer, and Allen W Wood. *Critique of Pure Reason*. Cambridge: Cambridge University Press, 1998.

———. *Groundwork of the Metaphysics of Morals*, ed. and transl., Mary Gregor, Cambridge: Cambridge University Press, 1998.

Kierkegaard, Soren. *Fear and Trembling and The Sickness Unto Death*. Translated by Walter Lowie. New York: Anchor Books, 1954.

Kenny, Anthony. *A New History of Western Philosophy*. New York: Oxford University Press, 2010.

Lemos, John. "Self-Forming Acts and the Grounds of Responsibility." *Philosophia (Ramat Gan)* 43, no. 1 (2015): 135–46. <https://doi.org/10.1007/s11406-014-9561-7>.

Locke, John. *An Essay Concerning Human Understanding*. Oxford: Oxford University Press. 2008.

Long, A. A., and D. N. Sedley. *The Hellenistic Philosophers*. Vol. 1: Translations of the Principal Sources with Philosophical Commentary. Cambridge: Cambridge University Press. 1987.

Libet, Benjamin. "Do We Have Free Will?" In *The Oxford Handbook of Free Will*, edited by Robert Kane, 551-558. First Edition. New York: Oxford University Press, 2001.

Lindemann, Hilde. *Damaged Identities, Narrative Repair*. Ithaca: Cornell University Press, 2001.

———. "Holding on to Edmund: The Relational Work of Identity." In *Naturalized Bioethics: Toward Responsible Knowing and Practice*. Edited by Lindemann, Hilde, Marian Verkerk, and Margaret Urban Walker. Cambridge: Cambridge University Press, 2009.

List, Christian. *Why Free Will is Real*. Cambridge: Harvard University Press, 2019.

MacIntyre, Alasdair C. *After Virtue: A Study in Moral Theory* Notre Dame, Indiana: University of Notre Dame Press, 1981.

Mckenna, Michael. "Ultimacy and Sweet Jane." In *Essays on Free Will and Moral Responsibility*, edited by Daniel Cohen and Nick Trakakis, 186–208. Newcastle upon Tyne: Cambridge Scholars Publishing, 2008.

McCann, Hugh J. "Making Decisions." *Philosophical Issues* 22, no. 1 (2012): 246–63. <https://doi.org/10.1111/j.1533-6077.2012.00228.x>.

———. *The Works of Agency: On Human Action, Will, and Freedom*. Ithaca, N.Y: Cornell University Press, 1998.

Mele, Alfred R. *Free Will and Luck*. Oxford: Oxford University Press, 2006.

———. "Free Will and Luck: Compatibilism versus Incompatibilism." *The Monist* 103, No.3 (July 2020): 262-277.

- . *Free: Why Science Has Not Disproved Free Will*. New York: Oxford University Press, 2014.
- . “Kane, Luck, and the Significance of Free Will.” *Philosophical Explorations* 2, no. 2 (1999): 96–104. <https://doi.org/10.1080/13869799908520970>.
- . “Kane, Luck, and Control: Trying to Get by without Too Much Effort.” In *Libertarian Free Will: Contemporary Debates*. First Edition, edited by David Palmer, 37–51. New York: Oxford University Press, 2014.
- . “Libertarianism, Luck, and Control.” *Pacific Philosophical Quarterly* 86, no.3 (August 2005): 381–407.
- . *Manipulated Agents: A Window to Moral Responsibility*. New York: Oxford University Press, 2019.
- . “Two Libertarian Theories: Or Why Event-Causal Libertarians Should Prefer My Daring Libertarian View to Robert Kane’s View.” *Royal Institute of Philosophy Supplement* 80, (May 2017): 49–68. <https://doi.org/10.1017/S1358246117000108>.
- . “Ultimate Responsibility and Dumb Luck.” *Social Philosophy & Policy* 16, no. 2 (1999): 274–93. <https://doi.org/10.1017/S0265052500002478>.
- Mill, John Stuart. *On Liberty*. Edited by Currin V. Shields. Indianapolis: Bobbs-Merrill, 1956.
- Nagel, Thomas. *Mortal Questions*. First Edition. Cambridge: Cambridge University Press, 1979.
- Nelkin, Dana K. “Moral Luck.” In *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta and Uri Nodelman. Stanford: Metaphysics Research Lab, Stanford University, 2023.
- Nelson, Katherine. “Narrative and the Emergence of a Consciousness of Self.” In *Narrative and Consciousness*. New York: Oxford University Press, 2003. <https://doi.org/10.1093/acprof:oso/9780195140057.003.0002>.
- Nietzsche, Friedrich. *Beyond Good and Evil*. New York: Penguin, 2003.
- Nichols, Shaun, and Joshua Knobe. “Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions.” *Noûs* 41, no. 4 (2007): 663–85. <https://doi.org/10.1111/j.1468-0068.2007.00666.x>.
- O’Connor, Timothy. “Agent-Causal Power.” In *Free Will*. Second Edition. Edited by Derk Pereboom, 308–328. Indianapolis: Hackett Publishing, 2009.
- . “Agent-Causal Theories of Freedom.” In *The Oxford Handbook of Free Will*. Second Edition. Edited by Robert Kane, 309–328. New York: Oxford University Press, 2011.

- . “Free Will and Metaphysics” In *Libertarian Free Will: Contemporary Debates*. First Edition. Edited by David Palmer, 27-36. New York: Oxford University Press, 2014.
- . *Persons and Causes*. Oxford: Oxford University Press, 2001.
- O’Shaughnessy, Brian. *The Will: A Dual Aspect Theory*. Cambridge: Cambridge University Press, 1980.
- O’Connor, Timothy. “Agent-Causal Theories” In *The Oxford Handbook of Free Will*, edited by Robert Kane, Second Edition. New York: Oxford University Press, 2011.
- Palmer, David. “Free Will and Control: A Noncausal Approach.” *Synthese* 198, no. 10 (2021): 10043–62. <https://doi.org/10.1007/s11229-020-02701-4>.
- . *Libertarian Free Will: Contemporary Debates*. First Edition. New York: Oxford University Press, 2014.
- . “Free Will, Libertarianism, and Kane.” In *Libertarian Free Will: Contemporary Debates*. First Edition. Edited by David Palmer, 3-14. New York: Oxford University Press, 2014.
- Paul, L. A. *Transformative Experience*. 1st ed. Oxford: Oxford University Press, 2014. <https://doi.org/10.1093/acprof:oso/9780198717959.001.0001>.
- Pereboom, Derk. *Free Will*. Second Edition. Indianapolis: Hackett Publishing, 2009.
- . *Living Without Free Will*. Cambridge: Cambridge University Press, 2001.
- . “The Disappearing Agent Objection to Event-Causal Libertarianism.” *Philosophical Studies* 169, no. 1 (2014): 59–69. <https://doi.org/10.1007/s11098-012-9899-2>.
- Reid, Thomas. “*Essays on the Active Powers of the Human Mind*.” In *The Works of Thomas Reid, D.D.* Volume 1. Third Edition. Edited by Sir William Hamilton. Edinburgh: MacLachlan and Stewart, 1852.
- Ricoeur, Paul. *Oneself as Another*. Chicago: University of Chicago Press, 1992.
- Rudd, Anthony. “In Defence of Narrative.” *European Journal of Philosophy* 17, no. 1 (2009): 60–75. <https://doi.org/10.1111/j.1468-0378.2007.00272.x>.
- . “Kierkegaard, MacIntyre and Narrative Unity-Reply to Lippitt.” *Inquiry* 50, no. 5 (2007): 541–49. <https://doi.org/10.1080/00201740701612416>.
- Rumi, Jalāl al-Dīn. *Rumi, Poet and Mystic*. Translated and edited by Reynold Alleyne Nicholson. London: Allen & Unwin, 1956.
- Sartre, Jean-Paul. *Being and Nothingness*. Translated and edited by Hazel Estella Barnes. New York: Washington Square Press, 1992.
- Sartorio, Carolina. “Free Will and Determinism: A Compatibilism.” In *Do We Have Free Will?*, 1st ed., 68–118. Routledge, 2022. <https://doi.org/10.4324/9781003212171-3>.

Scanlan, James P. "John Stuart Mill and the Definition of Freedom." *Ethics* 68 (1958): 194-.

Schechtman, Marya. *Staying Alive: Personal Identity, Practical Concerns, and the Unity of a Life*. First edition. Oxford: Oxford University Press, 2014.

———. "Self-Expression and Self-Control." *Ratio (Oxford)* 17, no. 4 (2004): 409–27. <https://doi.org/10.1111/j.1467-9329.2004.00263.x>.

———. "The Narrative Self." In *The Oxford Handbook of the Self*, edited by Shaun Gallagher, 394–416. Oxford: Oxford University Press, 2011.

———. *The Constitution of Selves*. Ithaca, NY: Cornell University Press, 1996.

———. "Art Imitating Life Imitating Art: Literary Narrative and Autobiographical Narrative." In *The Philosophy of Autobiography*, edited by Christopher Cowley, 22–38. Chicago: University of Chicago Press, 2015.

David Shatz, "Free Will and the Structure of Motivation." *Midwest Studies in Philosophy* 10, no. 1 (1986): 451–82. <https://doi.org/10.1111/j.1475-4975.1987.tb00551.x>.

Shabo, Seth. "More Trouble With Tracing." *Erkenntnis* 80, no.5 (November 2014): 987–1011.

Smith, Angela M. "Attitudes, Tracing, and Control." *Journal of Applied Philosophy* 32, no. 2 (2015): 115–32. <https://doi.org/10.1111/japp.12107>.

Smilansky, Saul. "Compatibilism: The Argument from Shallowness." *Philosophical Studies* 115, no. 3 (2003): 257–82. <https://doi.org/10.1023/A:1025146022431>.

———. *Free Will and Illusion*. Oxford: Clarendon Press, 2000.

———. "Free Will, Fundamental Dualism, and The Centrality of Illusionism." In *The Oxford Handbook of Free Will*. Second Edition. Edited by Robert Kane, 425–441. New York: Oxford University Press, 2011.

Sneddon, Andrew. "Alternative Motivation and Lies" *Analysis* 81, no 1 (January 2021): 46–52.

———. "Does Philosophy of Action Rest on A Mistake?" *Metaphilosophy* 32, no.5 (January 2003): 502–522.

Steward, Helen. *A Metaphysics for Freedom*. Oxford: Oxford University Press, 2012.

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

- . “The Impossibility of Moral Responsibility.” *Philosophical Studies* 75, no. ½ (1994): 5–24. <https://doi.org/10.1007/BF00989879>.
- Strawson, P. F. *Freedom and Resentment and Other Essays*. London ; Routledge, 2008. <https://doi.org/10.4324/9780203882566>.
- Taylor, Richard. *Action and Purpose*. Englewood Cliffs, New Jersey: Prentice-Hall, 1966.
- Taylor, Charles. *Sources of the Self: The Making of the Modern Identity*. Cambridge, Mass: Harvard University Press, 1989.
- Timpe, Kevin. *Free Will: Sourcehood and Its Alternatives*. London: Continuum Books, 2008.
- Ullmann-Margalit, Edna. “Big Decisions: Opting, Converting, Drifting.” *Royal Institute of Philosophy Supplement* 58 (2006): 157–72. <https://doi.org/10.1017/S1358246106058085>.
- Van Inwagen, Peter. *An Essay on Free Will*. Oxford: Clarendon Press, 1983.
- Vargas, Manuel. “The Trouble with Tracing.” *Midwest Studies in Philosophy* 29, no. 1 (2005): 269–91. <https://doi.org/10.1111/j.1475-4975.2005.00117.x>.
- Velleman, J. David. 2008. “The Way of the Wanton.” In *Practical Identity and Narrative Agency*, edited by C. Mackenzie and K. Atkins. New York: Routledge, 2008.
- Velleman, J. David. “What Happens When Someone Acts?” *Mind* 101, no. 403 (1992): 461–81. <https://doi.org/10.1093/mind/101.403.461>.
- . “The Self as Narrator.” In *Self to Self: Selected Essays*. Cambridge: Cambridge University Press, 2006.
- Watson, Gary. “Free Action and Free Will.” *Mind* 96, no. 382 (1987): 145–72.
- . “Free Agency.” *The Journal of Philosophy* 72, no. 8 (1975): 205–20. <https://doi.org/10.2307/2024703>
- . “Responsibility and the Limits of Evil: Variations on a Strawsonian Theme.” In *Free Will*, edited by Robert Kane, 164–188. Malden, Massachusetts: Blackwell Publishers, 2002.
- . “Soft Libertarianism and Hard Compatibilism.” *Journal of Ethics* 3 (1999): 353–368.
- Wegner, Daniel M. *The Illusion of Conscious Will*. Cambridge, Massachusetts: MIT Press, 2002.
- Williams, Bernard. *Moral Luck: Philosophical Papers 1973–1980*. Cambridge: Cambridge University Press, 1981.

Wolf, Susan. "Sanity and the Metaphysics of Responsibility." In *Free Will*, edited by Robert Kane, 164-188. Malden, Massachusetts: Blackwell Publishers, 2002.

Zemeckis, Robert. *Forest Gump*. Hollywood, CA: Paramount Pictures, 1994.

Zimmerman, David. "Hierarchical Motivation and the Freedom of the Will." *Pacific Philosophical Quarterly* 62, no. 4 (1981): 354–68. <https://doi.org/10.1111/j.1468-0114.1981.tb00073.x>.