DIMENSIONS OF RESPONSIBILITY: FREEDOM OF ACTION AND FREEDOM OF WILL*

By Robert Kane

Abstract: *In this essay, I distinguish two dimensions of responsibility: (i) responsibility for expressing the will (character, motives, and purposes) one has in action (voluntarily and without constraint) and (ii) responsibility for having the will one expresses in action. I argue that taking both of these dimensions into account is necessary to do full justice to our understanding of moral responsibility and our ordinary practices of holding persons responsible in moral and legal contexts. I further argue that the distinction between these dimensions of responsibility is importantly related to understanding age-old debates about the freedom of the will. For the first dimension of responsibility is historically related to the freedom of action—the power to freely express the will one already has in action. While the second dimension is historically related to the freedom of the will—the power to freely form or shape that will one may later express in action. And I argue that while the freedom of action so defined may be compatible with determinism, the freedom of will, and the deeper responsibility associated with it for forming one's own will, which I call "ultimate responsibility," are not compatible with a thoroughgoing determinism. In arguing throughout the essay for these claims and for the need to take into account both of these dimensions to do full justice to our understanding of moral responsibility, I consider ordinary practices of holding persons responsible in a variety of moral and legal contexts, discussing in the process H. L. A. Hart's "fair opportunity to avoid wrongdoing" criterion for assessing responsibility and blame in legal and criminal contexts, the relevance of recent experimental studies about folk intuitions concerning assessments of responsibility and blame, Harry Frankfurt's critique of the "principle of alternative possibilities," the distinction between "will-settled" and "will-setting" actions, and contemporary critiques of the very possibility and intelligibility of an ultimate responsibility for forming one's own will that would be incompatible with determinism.*

KEY WORDS: moral responsibility, blame, character, will, self-forming actions, free will

I. Introduction

I am pleased to be a contributor to this volume with so many able younger philosophers, some of whose work I am already familiar with and have learned from, and others I value the opportunity to learn from. That said, there are many disagreements between myself and the other contributors about critical issues regarding moral responsibility, free will,

* I owe a debt to many individuals whose comments and questions on this essay and other discussions helped me improve the final version in numerous ways, including among others, David Keyt, Keith Lehrer, Chandra Sripada, David Shoemaker, Justin Capes, Christopher Franklin, Carolina Sartorio, Jenann Ishmael, Fred Miller, Santiago Amaya, Dana Nelkin, and Elinor Mason. A special thanks is due to David Palmer and Michael McKenna whose written comments on the essay helped me clarify many points.

doi:10.1017/S0265052519000232

and related topics, that I will be discussing in this essay. Such disagreement does not trouble me, because encountering and addressing disagreements is essential to any progress in philosophy and has been essential to the development of my own views.

Plato intimated centuries ago that the Truth (with a capital T, one of his ideal Forms along with Justice and the Good) is not something any one person or group can wholly own, hoarding it from others like a pot of gold. The Truth in this ideal sense is something we can only *participate* in with other persons to some degree from our own limited or finite perspectives. That is why the pursuit of it is said to be a kind of love, a love of wisdom.

What I thought I would do therefore in this essay is discuss personally my own odyssey of thought about these matters, including personal experiences and encounters that have influenced it, and why I hold what I do. I will try to do this without losing rigor, but while also keeping you interested.

II. Dimensions of Responsibility

Let me begin with the two dimensions of responsibility of my title. I believe that if we are to do full justice to our understanding of moral responsibility and our practices of holding persons responsible, we must distinguish two dimensions of responsibility. Both are necessary for a fully adequate account of that understanding and those practices, and neither alone is sufficient. The first dimension is responsibility for *expressing the will* (the 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).

The second dimension of responsibility is another matter. It is not *responsibility* for *expressing* the will one *has* in action, but rather *responsibility* for *having* the will one *expresses* in action. 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 and intentionally in the first dimension, it was possible for you to have *also 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 the time. To be responsible in this second dimension, in other words, it cannot be *at all times* in one's life that only one possible action is determined by, and expressive of, your *already existing will*. Some choices or actions in one's life must be, as I like to say, *will-setting*, and not already *will-settled.* As is well-known, I call such will-setting actions "self-forming actions (SFAs)."

This distinction between two dimensions of responsibility is not the same as Gary Watson's well-known distinction between two "faces" of responsibility in his important article, "Responsibility and the Limits

of Evil."[1] But the distinctions are related. Watson distinguishes between responsibility as *attributability* and as *accountability*. By contrast, each of the two dimensions of responsibility defined here involve both attributability and accountability in his sense. For agents to be responsible for expressing the wills they have in action, as well as for having the wills they express in action, their actions and wills must be attributable to them and they must be accountable to some degree for them.

Each of the two dimensions also involves a third notion that plays a significant role in contemporary debates about responsibility, namely *answerability*.[2] Persons can be answerable to others for expressing the wills they have in action, but they can also be answerable for *having* the wills they express in action, and hence for *being* the sorts of persons they are with the "ill wills" or "good wills" their actions toward others express. Many philosophers speak here about the "quality of will" that agents express in their actions. Whether agents are answerable and accountable for the quality of the wills they express in action, whether for ill or for good, as well as for the actions themselves that are the products of their wills, has always seemed to me a significant matter if we are looking for a complete account of moral responsibility. Hence the emphasis on the two dimensions. Aristotle was on to the idea when he suggested that if a man is responsible for the wicked acts that flow from his character, he must also be responsible for the wicked character from which these acts flow.[3]

Of further importance here is that these distinctions are crucially related, in my view, to adequately understanding age-old debates about the freedom of the will. For the first of these dimensions of responsibility is historically associated with the *freedom of action*—the ability to freely express the will you already have in action. Whereas, the second dimension of responsibility is related to *freedom of will*—the freedom to *form* or *shape* the will that you will express in action. Much modern philosophy, from the seventeenth century and into the twentieth and twenty-first

[1] Gary Watson, "Responsibility and the Limits of Evil: Variations on a Strawsonian Theme," in F. D. Schoeman, ed., *Responsibility, Character, and the Emotions* (Cambridge: Cambridge University Press, 1987), 256–86.

[2] David Shoemaker has written extensively and astutely about the relation of these three notions of responsibility—attributability, accountability, and answerability—spelling out their implications and relations to one another in enlightening detail. See his *Responsibility From the Margins* (New York: Oxford University Press, 2015).

[3] Aristotle, *Nichomachean Ethics*. Vol. 9 of *The Works of Aristotle*, W. D. Ross, ed. (London: Oxford University Press, 1915), 1114a13–22, 255a8, 110a17, 1113b21, 1114a18–19. In comments on this essay, both Fred Miller and David Keyt pointed out the similarities between much that I say about responsibility for one's will and what Aristotle says about responsibility for character. His view, they argue, is that character is formed by the kinds of actions we take. Vicious actions create and sustain a vicious character and virtuous actions a virtuous character. I agree, but would add that character traits are only one (albeit important) aspect of what is historically designated as the will, which includes not only persistent traits of character, but also motives and preferences (which may be long or shorter term) and intentions, which are the immediate products of choices. The will from which we act includes all of these.

centuries has attempted to reduce the "problem of free will" to a problem of "freedom of action," thereby in my view obscuring the traditional problem of free *will* and making it appear simpler than it is. Free will is not just about free action, though it involves free action. It is about *self-formation*, about the formation of our "wills" or how we got to be the kinds of persons we are with the characters, motives, and purposes we now have; and this is distinctly related to the second dimension of responsibility, which I call "*ultimate responsibility* (UR)." Are *we* ultimately responsible to some degree for having the wills we do have, or can the sources of our wills be completely traced backward to something over which we had no control, such as Fate or the decrees of God, heredity and environment, social conditioning, and so on? Therein, I believe, lies the core of the traditional problem of "free *will*."

John Locke was one of the first to start this modern trend in the seventeenth century of reducing the problem of free will to one of free action. Locke famously wrote that the problem is not about freedom of will, but rather about the freedom of the agent or *free agency*.[4] Like other thinkers of the era and since, Locke was skeptical of medieval notions of the will that often reduced it to an obscure inner homunculus with magical powers. He believed medieval notions of will were outdated, had no place in the new sciences of the day, and should go the way of witches and phlogiston. Now my view is that Locke was right in saying that the traditional problem of free will *was* about free agency. But where he went wrong, in my view, is in failing to note that there are multiple dimensions of *free agency*. And one of these dimensions, freedom of will, is the particular source of traditional concerns and historical debates about *free will* and determinism. For freedom of action, as I see it, in the sense defined here, the freedom to express your will in action, voluntarily and without constraint, may well be compatible with determinism. But freedom of will is not, for reasons I will shortly be discussing.

III. Walden Revisited

But why do we need both dimensions of freedom to give a full accounting of freedom and, ultimately, of responsibility as well? I'll tell a brief story here. When quite young, I don't remember exactly when, I read B. F. Skinner's utopian novel *Walden Two*.[5] And I recall Frazier,

[4] John Locke, *An Essay Concerning Human Understanding*, P. Nidditch, ed. (Oxford: Oxford University Press, 1975), Bk. II, chap. xxi, 134. Locke himself seems to qualify his position in a later edition of this work, in what I would regard as the fruitful direction, when he says: "Yet there is a case wherein a Man is at Liberty in respect of willing; and that is the choosing of a remote Good as an end to be pursued. Here a Man may suspend the act of his choice from being determined for or against the thing proposed, till he has examined whether it be really of a nature, in itself and consequences, to make him happy, or no" (Ibid., 139).

[5] B. F. Skinner, *Walden Two* (New York: MacMillan, 1962).

one of the designers of the community, boasting to the visitors he was showing around the place that Walden Two was the freest place on earth, the freest place imaginable. There were no jails, no need for them, no restraints. People could have and do whatever they wanted or chose, he said, because they were conditioned from childhood to want and choose only what they could have and do. And my reaction was: "What? This is the freest place imaginable?" Something seems to be missing here. And this was a common reaction among my schoolmates. For in one sense there was a great deal of freedom in Walden Two, but in another sense it was a *controlled* society.

My youthful interest in *Walden Two* was prompted in part by the fact that where I grew up was not too far from the original Walden. It was, of course, a pond, not a society, which as high schoolers, we occasionally visited on warm summer evenings. I asked myself at the time what Henry David Thoreau would have thought about Skinner's *Walden Two*. Would Thoreau have thought it to be the freest place imaginable? And the answer seemed to be clearly "no." He would not.

I lacked the resources then to say in more detail what seemed to be missing in *Walden Two*—to say, as I eventually would, that there were different dimensions of freedom (a complex notion, to be sure). And what Frazier and his colleagues were doing was maximizing the freedom in one dimension—surface freedom of action—by minimizing it in another dimension, the freedom of agents to form their own wills and what sorts of persons they would become—a deeper freedom of the will. If freedom is not a one-dimensional concept, but something with multiple dimensions, one can maximize it in some dimensions while minimizing it in others. And freedom *is* a complex notion. You can have it to varying degrees in some dimensions, but not in others; and there necessarily are trade-offs. Political theorists know this all too well. And responsibility is correspondingly complex.

Now, let's be clear here, Skinner was writing in the heyday of behaviorism, which soon after became passé. There are plenty of legitimate doubts that one could exercise such control over human behavior by behavioral conditioning alone. If such control were possible at all, most would now believe it would have to involve genetic and neurological conditioning of humans from earliest ages. But the same dialectic would hold, whatever the means—*if* people were completely conditioned from childhood so that they would *always* want and choose only what they could have and do. Note here that worries about this are greater *to the extent* that the conditioning *by whatever means* implies that the agents could *never* in the course of their lives *voluntarily and intentionally choose or do otherwise* than they were conditioned to do—that they could *never resist this conditioning*. In other words, the worry grows to the extent that the conditioning may be completely *determining*, whatever form the conditioning may take.

In fact, it was intimated in the novel that some young people might chafe under the restrictions on what they could have or do in *Walden Two* to the degree that the founders' conditioning might not be completely successful. These young people would not necessarily think the Walden Two way of life was bad. But they would want to chart their own course, make their own decisions about what they might want or choose to become, even if that meant taking chances of choosing wrongly and being disappointed, or worse. In other words, they wanted the experience of *self*-formation, and were willing to *take responsibility* for what the consequences of their life-forming actions should be. Anyone who has raised teenagers knows what I mean. They may respect you and what you are, but they do not want to be you. They want to be themselves.

IV. Compatibilism and Incompatibilism about Responsibility and Freedom

But I said I would be returning to the reasons why I believe that, while the first dimension of responsibility and the freedom of action related to it may be compatible with determinism, the second dimension of responsibility for forming oneself and the freedom of will it involves, are incompatible with a thoroughgoing determinism. It will be helpful to begin by taking note of a number of bad arguments for the supposed incompatibility of responsibility and determinism that have been given by ordinary folk and even some philosophers, that play no role in my own thinking. Some of these arguments have appeared in recent developments in experimental philosophy, involving experiments to determine what ordinary or folk intuitions may be about whether or not moral responsibility and freedom are compatible with determinism. Many of these experiments, due to various philosophers, including prominently Eddy Nahmias, find, according to the experimenters, that the majority of the folk are intuitively compatibilists about responsibility and freedom. They believe responsibility and freedom are compatible with determinism, thus taking issue with both libertarians about free will and moral responsibility, like myself, and free will skeptics, who believe that free will and moral responsibility are not compatible with determinism.[6]

What is relevant for present purposes is that Nahmias and others also interestingly argue that the minority of respondents who do give *in*compatibilist responses to various deterministic scenarios in such studies are in error. They give such responses, it is said, because they erroneously believe that determinism involves "bypassing" the mental. That is, they believe that if an agent's actions were determined, their actions would

[6] For an overview of this work, see Eddy Nahmias, "Intuitions about Free Will, Determinism and Bypassing," in Robert Kane, ed., *The Oxford Handbook of Free Will*, 2nd ed. (Oxford: Oxford University Press, 2011): 555–76.

have been caused in such a way that their mental states and processes — beliefs, desires, and reasonings—would be bypassed and would not make a difference to what they end up doing. And this is an error, they argue, that leads these folk to believe mistakenly that moral responsibility and free will are incompatible with determinism.

To complicate matters, in an important recent book, Shaun Nichols, argues that many other studies in experimental philosophy, conducted by himself and others, show contra Nahmias and other theorists that the majority of folk give *incompatibilist* responses when presented with deterministic scenarios, suggesting that they believe moral responsibility and free will are not compatible with determinism.[7] But while taking issue with Nahmias and others on these results, Nichols interestingly argues that those in the majority in the studies he cites who give incompatibilist responses are also guilty of an error, though a different error. The error is not that they believe determinism bypasses the mental, but rather that they believe their conscious experiences of freely choosing and acting provide convincing evidence that their choices and actions could not be determined. And this too is an error, Nichols argues.

So we have mixed results in these studies concerning whether folk intuitions about moral responsibility and free will are compatibilist or incompatibilist. But we have agreement by the *experimenters* that those who give incompatibilist responses, whether they are in the majority or minority, are guilty of some sort of philosophical error. Now I agree that the "errors" Nahmias and Nichols describe are indeed errors. While I have always had strong incompatibilist intuitions about free will and ultimate responsibility, these intuitions were never remotely based on beliefs about bypassing the mental or beliefs that our ordinary experience of choosing and acting provided reliable evidence against determinism. To the contrary, if anyone of the folk, or any philosophers for that matter, had expressed such beliefs to me, I would have told them straight off that they were in error.

They were in error, I would have said, because the free will problem has arisen in history whenever persons have wondered whether their choices and actions might be determined by factors and circumstances *unknown* to them and beyond their control. In short, *the very problem arises when one realizes that our ordinary experience of choosing and acting does not provide reliable evidence against determinism*. As for bypassing, I would have said that it is a mistake to suppose determinism implies bypassing of the mental. Determinism is the view that given the past at any time, there is only one possible future. And this means the *whole* past. No existing events or states, physical *or mental*, are excluded from it. To exclude the agent's psychological history from the whole past when defining determinism is

[7] Shaun Nichols, *Bound: Essays on Free Will and Responsibility* (Oxford: Oxford University Press, 2015).

to solve the free will problem by fiat rather than by honest toil. In my very first book, written in the early 1980s, I made these points a central theme and challenge.[8]

On what, then, were my long-held incompatibilist intuitions about free will and ultimate responsibility based? I will focus here on causal determinism, to keep matters simple, though what I say could be generalized. To put the matter succinctly, if determinism was true, anything you might have done differently in the course of your life to make yourself different than you are would have been *causally impossible*.[9] Your character, your motives, your dispositions, your intentions, the quality of your will, at any time—anything you might have done to make anything different about yourself at any time—would have been causally impossible. And serious questions arise about whether persons can be morally obligated to do what it is causally impossible for them to do or blamed for failing to avoid doing something that it was causally impossible for them to avoid doing.

Now it might be objected that the causal impossibility of doing something or doing otherwise is not among the usual excusing or exempting conditions commonly *cited* in ordinary practices of holding persons responsible—such as incapacity, ignorance, lack of opportunity, coercion, duress, insanity, and mental impairment—that compatibilists rightly insist upon. But it does not follow that the causal impossibility of doing something or doing otherwise is not often itself an excusing condition or that it does not often *imply* one or another of these commonly cited excusing conditions, such as *lack of opportunity to exercise a capacity in particular circumstances*. If an airplane runs off a runway, injuring many passengers, the pilot may initially be blamed for negligence or incompetence in failing to avert the accident. But if further investigation shows there was a subtle defect in the design of the aircraft that made it *causally impossible* for the pilot to have avoided it, despite his efforts, he would rightly be excused of responsibility and blame for the accident. Similarly, if an explosion occurs in a factory, the guard in the control room in charge of monitoring conditions may be legitimately excused of responsibility or blame for not preventing the explosion if an investigation shows that it was causally impossible for her to have done so because of a defect in

[8] Robert Kane, *Free Will and Values* (Albany, NY: State University of New York Press, 1985).

[9] To be more precise about what is meant here and throughout by "causal impossibility": The occurrence of a state or event, E, at a time, t, (E_t), is *causally impossible*, if and only if, there are states and/or events occurring prior to t (P_{t-}) and laws of nature (L), such that <It is not possible that (P_{t-} & L & E_t)>. The laws involved in this definition, being natural laws, need not be logically necessary. They need not hold in all logically possible worlds. What is required, however, if E_t is causally impossible is that in every logically possible world in which the laws L do hold and the relevant states and/or events, P_{t-}, occur prior to t, E does not occur at t. Its occurring, *given* these laws and these prior states and events, is logically impossible. I am indebted here to David Palmer, whose comments on the essay impressed on me the importance of spelling out the meaning of "causal impossibility" employed in it in more precise terms.

the monitoring systems. These agents had certain *general* capacities, but it was causally impossible for them to *exercise* these capacities *in the circumstances*; and so they are excused of responsibility and blame.

Such examples, which could be multiplied indefinitely, show that the causal impossibility of doing something or doing otherwise *is* relevant to ordinary practices of holding persons responsible for what they have done. But it is equally important to point out that causal impossibility and the consequent lack of opportunity to do something or do otherwise in certain circumstances, do not always imply that agents are excused from moral responsibility. They imply this only *conditionally*. And this is an equally crucial fact about the nature of moral responsibility and its relation to free will.

If, for example, it could be shown that it was causally impossible for a drunk driver to have avoided hitting a pedestrian on a dark and rainy night, given all the circumstances at the moment of the accident, that fact alone will not excuse him of responsibility. For one must also ask whether he was responsible by virtue of earlier actions or omissions for the existence of some of those crucial circumstances that made it now causally impossible for him to avoid the accident, such as prior decisions to drink and then drive. The causal impossibility of doing otherwise now will not constitute an excuse, if some of the crucial circumstances which make it now causally impossible to do otherwise were the results of actions or omissions by the agent in the past, *which it was not causally impossible for the agent to have avoided when they occurred.* And this last phrase, which I have italicized, is crucial.[10] For the problem is that if determinism is true, there would be no actions or omissions in an agent's past that were not causally impossible for the agent to have avoided when *they* occurred.

Ah, but a common response at this point is that if the causal impossibility of now doing otherwise is due to the *quality of will* of the agents when they act, whether they act from an ill will or good will, that is different than if the causal impossibility is due to external factors over which they have no control and are not responsible (such as faulty aircraft design or monitoring systems).

[10] As David Palmer correctly points out in written comments on this essay, many writers on moral responsibility speak here of a distinction between *indirect* or derivative responsibility and *direct* or nonderivative *responsibility*. The drunk driver is indirectly or derivatively responsible for killing the pedestrian in this example by virtue of earlier actions, such as drinking and driving, for which he is directly responsible. I accept such a distinction, but would spell it out somewhat differently in ways that sometimes depart from the way some writers distinguish indirect and direct responsibility. I distinguish two kinds of responsible action that are done "of an agent's own free will." The first are actions done of an agent's own free will in the sense of a will that was formed by the agent by prior self-forming actions (SFAs). The second are the self-forming actions or SFAs themselves by which the agent forms the will from which the agent subsequently acts. These two kinds of responsible action are related to what many call indirectly and directly responsible actions, respectively, though there are some nuanced differences. What they do correspond to are the two dimensions of responsibility of the title of this essay: responsibility for expressing the will one has an action and responsibility for having the will one expresses in action. It is beyond the scope of this essay to discuss the nuanced differences between these distinctions and the distinction between what many call indirect and directly responsible actions. But the distinctions are related, if not identical.

But if agents are to be responsible for the present quality of their wills when they act, there must have been some acts in their prior histories by which the present quality of their wills were formed, *which were not causally impossible for the agents to have avoided when they occurred.* And if determinism was true, there could have been no such acts now or at any time in the past.

V. Responsibility and Blame in Criminal Contexts

These points can be further illustrated by considering cases of assessing responsibility and blame in criminal proceedings. A well-known example in the literature of this kind is Gary Watson's searing account of the ruthless murderer, Robert Harris, on death row in California for multiple murders.[11] Another is an example that I have used in my writings, roughly based on personal experience, triggered by the trial of a young man accused of assault and rape.[12] I will discuss it here, since it brings out some crucial points Watson does not highlight.

My initial reactions attending the trial of this young man were filled with anger and resentment against him, since we knew the family of the teenage girl who was the victim and lived in our neighborhood. But as I listened daily to the testimony of how the young man came to have the mean character and perverse motives he did have—a sordid story of parental rejection, sexual abuse, bad role models, and other factors (not entirely unlike the case of Robert Harris)—some of my resentment toward the young man decreased and was directed toward other persons who abused and influenced him. But—and here is a key point—I wasn't yet ready to shift all the blame away from the young man himself. I resisted this "transference of responsibility" and "blame" entirely to others and wondered whether some residual responsibility might not belong to him. My question became: Was his behavior *all* a question of bad parenting, neglect and abuse, social conditioning, and like factors, or did he have any role to play in choosing it that was not made inevitable by these factors?

We know that parenting and society, genetic makeup and upbringing, have a profound influence on what we become and what we are. But were these influences entirely *determining*, or did they "leave anything over" for the young man to be responsible for? Note that the question of whether he was merely a victim of bad circumstances or had some residual responsibility for being the way he is—the question, that is, of whether he became the person he is with the will he has to any degree *of his own free will*—seems to depend on whether these other factors were or were not *entirely* determining. It seems to depend, in other words, on whether or not it was *ever causally possible for him to have resisted the influences of his past*

[11] Watson, "Responsibility and the Limits of Evil," passim.

[12] E.g., in Robert Kane, *A Contemporary Introduction to Free Will* (Oxford: Oxford University Press, 2005), 4–5.

upbringing and to have acted differently at some points in the course his lifetime *to make himself different than he now is*. And if determinism is true, resisting these influences by acting differently than he actually did at *any* time in his lifetime would have been *causally impossible*.

Now one might further argue that my particular reactions at this trial to the young man, the fact that my reactive attitudes of resentment and blame toward him were mitigated to some degree and transferred to others when I learned about his sad history, were the reactions of a "philosopher" and not the reactions of ordinary folk. But this was far from being the case. My wife and I sat in this courtroom with friends and other neighbors of the young girl's family, none of whom were philosophers. They were firemen, businesswomen, store owners, high school football coaches, teachers, and many others; and all had similar reactions to ours. Keep in mind that, like us, they all resisted mightily transferring responsibility entirely away from the young man. But their reactive attitudes, including retributive ones, were nonetheless mitigated to some degree and influenced by hearing the sordid stories of his history.

Moreover, if there *were* any persons in that courtroom whose retributive attitudes were in no way influenced by listening to the history of the young man (as I am sure there were), then I would not want to see them anywhere near a jury deciding the fate of persons I cared about, or any other persons whatever. For they would not be capable of responding in ways I believe would be *fair* to those they judge. They would not be capable of responding fairly, if they were not capable of appreciating that, to the extent that the young man's sad history made it *causally impossible* for him to have turned out differently, to that extent he would not have had a *"fair opportunity to avoid wrongdoing*."

VI. The "Fair Opportunity" Criterion and the Principle of Alternative Possibilities (PAP)

Legal theorist, H. L. A. Hart posited an influential "fair opportunity" criterion for responsibility in criminal contexts.[13] He argued that a necessary condition for ascribing responsibility and culpability to agents in such contexts is that the agents must have had a "fair opportunity to avoid wrongdoing," or more generally, a "fair opportunity to have done otherwise." In an important recent article, David Brink and Dana Nelkin argue that Hart's criterion is important for understanding moral responsibility generally as well as legal responsibility, and is a crucial part of what they call the "architecture" of ordinary practices of ascribing moral and legal responsibility.[14]

[13] H. L. A. Hart, *Punishment and Responsibility* (New York: Oxford University Press, 1968).

[14] David Brink and Dana Nelkin, "Fairness and the Architecture of Responsibility," in David Shoemaker, ed., *Oxford Studies in Agency and Responsibility*, Vol. 2 (New York: Oxford University Press, 2013), 283–87.

I agree. But my questions are these: *Do agents have a fair opportunity to do otherwise, if their doing otherwise was always causally impossible?* Did they have a fair opportunity to become different than they are, if anything they might have done in the course of their lives to make themselves different than they are was causally impossible?

Another important thing to note about Hart's influential criterion, not often noted, is that having a "fair opportunity to have done otherwise" in moral and legal contexts is not the same thing as "having the power at the time of one's actions to do otherwise" than one actually did. The drunk driver, for example, may be judged to have had a fair opportunity to have done otherwise than he did by virtue of actions taken at earlier times, even when he lacked the power and opportunity to have done otherwise at the time he hit the pedestrian.

There is a more general point implied here about moral responsibility that is crucially important and too often overlooked, a point that I have often emphasized in my writings. It is that being able "to do otherwise than you do," that is, the much discussed "*alternative possibilities*" or AP criterion for free action, is *neither a necessary nor a sufficient condition for being responsible in both dimensions* of responsibility. It is not necessary because agents can be responsible for expressing the wills they have and for having the wills they express in many instances in which they act without constraint, satisfying all of the usual compatibilist criteria for freedom and responsibility, and do so *of their own free will* in the sense of a will of their own prior free-making. In other words, they can be responsible in both dimensions even when they could not have willingly done otherwise, given the quality of their wills, at the time they acted—and even if their actions were determined by the quality of their wills (plus background circumstances) when they acted. This can be so to the extent that the present quality of their wills was the result of prior choices and actions in the course of their lives (self-forming actions or SFAs), for which they could have done otherwise in a manner that was not determined by their pre-existing will at the time they acted.

It follows that Harry Frankfurt's much-discussed Principle of Alternative Possibilities (PAP)—"persons can be morally responsible for their actions only if they could have done otherwise than perform them"—is false, as he claims.[15] Nor do we have to appeal to his much-discussed "Frankfurt-style examples" to establish this. There are more general reasons to believe that PAP is false, as just noted. But while it is not necessary to have the power to do otherwise with respect to an action in order to be responsible for it in both dimensions, it *is* necessary in order to be responsible in both dimensions that *some* acts performed in the course of one's life (self-forming actions, SFAs), by which one formed the will from which one later acts, be such that one had the power to do otherwise with

[15] Harry Frankfurt, "Alternate Possibilities and Moral Responsibility," *Journal of Philosophy* 66 (1969): 829–39.

respect them, in a manner not determined by one's pre-existing will plus background circumstances when they occurred. And I have argued in various writings that the kind of control exercised over agents in Frankfurt-style examples of all varieties, in deterministic or indeterministic settings, would make such self-forming actions impossible.[16]

The reasons have to do with the further significant fact: for self-forming actions, being able to do otherwise, though necessary, is not *sufficient* if they are to be genuinely *will-setting*, and not merely *will-settled*, even if they are also *undetermined*. For actions to be *will-setting*, and not merely will-settled, agents must not only have the powers to perform them and do otherwise. They must have the power to perform them *voluntarily* and *intentionally*, that is, *willingly*, and at the same time the power to do otherwise *voluntarily* and *intentionally—either way* "at will," as we say, rather then being able to do otherwise only inadvertently, involuntarily, unintentionally, by accident or mistake. What is required, in other words, is for agents to have what I call *plural voluntary control* (PVC) over their self-forming actions. And in Frankfurt-style examples of all kinds, *even when the controllers do not actually intervene*, their mere presence (given their powers and intentions) makes it impossible for agents to *willingly* do otherwise than the controllers wish.[17]

[16] E.g., Robert Kane, "Responsibility, Indeterminism and Frankfurt-style Cases," in David Widerker and Michael McKenna, eds., *Moral Responsibility and Alternative Possibilities* (Aldershot, UK: Ashgate, 2003); Robert Kane, "Frankfurt-Style Examples and Self-forming Actions," in Ishtiyaque Haji and Justin Caouette, eds., *Free Will and Moral Responsibility* (Newcastle upon Tyne: Cambridge Scholars Publishing, 2013), 58–73.

[17] In comments on this essay, Michael McKenna correctly points out that this argument against FSEs departs in some important respects from earlier arguments I have made against such examples. As early as my 1985 book (see note 8), I argued that the usual FSEs put forward by Frankfurt and many others illicitly presuppose determinism since the controllers require a reliable "prior sign" in order to know what the agent is about to do and hence whether or not to intervene. Such reliable prior signs, however, will not be available if, as libertarians about free will require, some free actions must be undetermined up to the moment they occur. Similar arguments against FSEs were later put forward by others, such as David Widerker, "Libertarianism and Frankfurt's Attack on the Principle of Alternative Possibilities," *Philosohical Review* 104 (1995): 247-61 and Carl Ginet, "In Defense of the Principle of Alternative Possibilities," *Philosophical Perspectives* 10 (1996): 403–17. In the decades following, however, various defenders of FSEs, such as Alfred Mele and David Robb, "Rescuing Frankfurt-style Examples," *Philosophical Review* 107 (1998): 97–112, Derk Pereboom, *Living Without Free Will* (Cambridge: Cambridge University Press, 2001), and others, produced new examples of FSEs designed to work in cases where the choices or actions to be controlled were undetermined, as libertarians require. In response to these new FSEs, I have argued (see note 16 for the relevant articles), as I do in this essay, that while these more recent FSEs do take into account the indeterminism required of libertarian free choices, they rule out another condition of equal importance that libertarian free choices which are SFAs, must satisfy. Namely, the presence of the controller in such indeterministic FSEs, whether or not the controller actually intervenes, makes it impossible for the agents to have plural voluntary control (PVC) over the controlled choices, i.e., the power at the time to willingly make them and the power to willingly do otherwise. Thus I argue, as I do here, that the kind of control exercised over agents in FSEs of all varieties, deterministic *or* indeterministic, would make such SFAs impossible. So that if libertarian free choices that are self-forming satisfy both conditions, as I believe they must, being undetermined and the agents having PVC over whether or not they occur, they cannot be controlled by Frankfurt controllers.

VII. Are Ultimate Responsibility and the Freedom of Will It Requires Intelligible and Possible?

Whatever the force of the preceding arguments may be, we know that a majority of philosophers and scientists in the modern era and especially in the past century have deep suspicions about any notions of responsibility and freedom that would be incompatible with determinism. Nietzsche summed up these suspicions in his inimitable prose when he said that "the desire for 'freedom of the will' in the superlative metaphysical sense . . . the desire to bear the ultimate responsibility for one's actions oneself . . . to be nothing less than a *causa sui* [cause of oneself] . . . is the best self-contradiction that has been conceived so far [by the human mind]."[18] Modern compatibilists and free will skeptics have consistently argued that if notions of ultimate responsibility and free will must be *incompatible* with determinism, they are not something we could have anyway.

The reasons go back to an ancient dilemma: if the freedom required for moral responsibility is *not* compatible with determinism, it would not be possible at all, since it could not be compatible with *indeterminism* either. Undetermined events, it has been argued since the time of the Stoics, would occur spontaneously and hence could not be controlled by agents in the way that *free* and *responsible* actions require. If, for example, a choice occurred by virtue of some undetermined (say, quantum) events in one's brain, it would seem a fluke or accident rather than a responsible choice. Such undetermined events in our brains or bodies would not enhance our freedom and control over, and hence responsibility for, our actions, but rather would diminish freedom, control, and responsibility.

Can one make sense of free choices or actions that are morally responsible and yet undetermined in a way that avoids their being reduced to matters of mere luck or chance, on the one hand, or to mystery, or what P. F. Strawson called the "panicky metaphysics of libertarianism" on the other?[19] Though it is fair to say that the majority view among contemporary philosophers and scientists is that this cannot be done, I believe it can be done and have attempted to do so in a series of writings over the past four decades.[20] What I learned was that no simple solution would

[18] Friedrich Nietzsche, *On the Genealogy of Morals*, trans. Walter Kaufmann (London: Vintage Books, 1989), Sec. 17.8, 89.

[19] P. F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962): 21.

[20] See, e.g., Kane, *Free Will and Values*; Robert Kane, *The Significance of Free Will* (Oxford: Oxford University Press, 1996); "Responsibility, Luck and Chance: Reflections on Free Will and Indeterminism," *Journal of Philosophy* 96, no. 5 (1998): 217–40; "Some Neglected Pathways in the Free Will Labyrinth," in *The Oxford Handbook of Free Will*, ed. Robert Kane (Oxford: Oxford University Press, 2002): 406–37; "Rethinking Free Will: New Perspectives on an Ancient Problem" in *The Oxford Handbook of Free Will*, 2nd edition, ed. Robert Kane (Oxford: Oxford University Press, 2011), 381–404; "New Arguments in Debates on Libertarian Free Will," in *Libertarian Free Will: Contemporary Debates*, ed. David Palmer (Oxford: Oxford University Press, 2014), 179–214.

be forthcoming, if it could be done at all. Rather a complex tapestry of ideas would be required.[21] While I cannot discuss this whole tapestry as I envisage it here, I would like to say enough to show how the ideas and distinctions discussed in this essay have an essential role to play in it, including the two dimensions of responsibility, the distinction between freedom of action and will, the role of "self-forming actions," the distinction between "will-setting" and "will-settled" actions, the notion of "fair opportunity to do otherwise," and others.

Consider self-forming actions: I argue that they occur at those times in life when we are torn between competing visions of what we should do or become; and they are more frequent in everyday life than one may think. Perhaps we are torn between doing the moral thing or acting from ambition, or between powerful present desires and long term goals, or faced with difficult tasks for which we have aversions. In such cases, we are faced with competing motivations and have to strive or make an effort to overcome the temptation to do something else we also strongly want. At such times, the tension and uncertainty we face about what to do, I suggest, would be reflected in some indeterminacy in our neural processes themselves (for example, in the form of nonlinearly amplified quantum fluctuations at the neuronal level)—"stirred up," one might say, by the conflicts in our wills. The uncertainty and inner tension we feel at such moments of self-formation would thereby be realized in some indeterminacy in our neural processes themselves. The experienced uncertainty would correspond physically to the opening of a window of opportunity that temporarily screens off complete determination by the past.

In such cases of self-formation, where we are faced with competing motivations, whichever choice we might make would require an effort

[21] Part of this task of making sense of such a free will and the associated tapestry of ideas involves considering the empirical and scientific question whether any indeterminism is there in the brain in ways appropriate for free will. No purely philosophical theory can settle this matter. It is interesting, however, that in the past decade there has been more openness and discussion on the part of scientists and philosophers about this possibility and it remains an open scientific question. See, e.g., Paul Glimcher, "Indeterminacy in Brain and Behavior," *Annual Review of Psychology* 56 (2005), 25–56; Robert C. Bishop, "Chaos, Indeterminism and Free Will," in *The Oxford Handbook of Free Will*, 2nd ed., Robert Kane, ed. (Oxford: Oxford University Press (2011), 84–100; B. Brembs, "Towards a Scientific Concept of Free Will as a Biological Trait," *Proceeding of the Royal Society B: Biological Sciences* 278 (2011): 930–39; Stuart Hameroff and Roger Penrose, "Conscious Events as Orchestrated Space-Time Selections," *Journal of Consciousness Studies* 3 (1996): 36–53; Peter Jedlicka, "Quantum Stochasticity and (the End of) Neurodeterminism," in Antonella Corradini and Uwe Meixner eds., *Quantum Physics Meets the Philosophy of Mind* (Berlin/Boston: Walter de Gruyter, 2014), 183–97; Martin Heisenberg, "The Origin of Freedom in Animal Behavior," in A. Suarez and P. Adams eds., *Is Science Compatible with Free Will?* (Berlin: Springer Verlag, 2013), 95–103; Peter Ulric Tse, *The Neural Basis of Free Will* (Cambridge, MA: MIT Press, 2013); Michael Shadlen, "Comments on Adina Roskies: Can Neurosciences Resolve Issues about Free Will?" in W. Sinnott-Armstrong ed., *Moral Psychology* Vol. 4 (Oxford: Oxford University Press, 2014): 175–87; Henry Stapp, *The Mindful Universe* (Berlin: Springer, 2007); Mark Balaguer, *Free Will as an Open Scientific Problem* (Cambridge, MA: MIT Press, 2010).

of will or *exercise of will power* to overcome the temptation to make the other choice. I thus postulate, in such cases that multiple goal-directed cognitive processes (volitional streams, as I call them) would be involved in the brain, corresponding to these exercises of will power, each with a different goal corresponding to the different choices that might be made—in short, a form of parallel processing in the free decision-making brain. One of these volitional processes would have as its goal making one of the competing choices (say, a moral choice), realized by reaching a certain activation threshold, while the other has as its goal the making of the other self-interested choice. And the processes would have different inputs—moral motives on the one hand, self-interested ones, on the other. In such circumstances, if either cognitive process succeeds in reaching its goal (the particular choice aimed at) despite the indeterminacy involved, the resulting choice would be brought about by the agent's effort or exercise of will power to bring about that choice for those motives; hence it would be brought about by the agent.

In such a picture, the indeterminism involved would not be a cause acting on its own, but an ingredient in these larger goal-directed cognitive activities of the agent, in which the indeterminism functions as an interfering element in the attainment of the goal, lowering the probability, without eliminating the possibility, that the goal will be attained. The choices that result would then be achievements brought about by the goal-directed activity (the striving or effort) of the agent, which might have failed since it was undetermined, but did not. Moreover, if there are multiple such processes aimed at different goals (in the conflicted circumstances of an SFA), whichever choice may be made will have been brought about by the agent's volitional striving or exercise of will power to make that particular choice rather than the other, despite the possibility of failure due to the indeterminism.[22] There are, of course, many questions and objections that need to be addressed about this view that I have attempted to address elsewhere and cannot explore further within the confines of this essay.

For, in conclusion, I want to emphasize that the amount of ultimate moral responsibility and freedom of will we might have on the view thus described—if such a view can be made to be coherent at all—would always

[22] One could imagine simpler cases in which will power is exercised only in one direction and the alternative choice is the result of "weakness of will." Such cases can occur. But the resulting choices would not be SFAs in the sense intended here because they would not be "will-setting" choices. The alternative choice would rather be the result of a failure to do what the will was "set" on doing (hence, "weakness of" will). In such cases, you would not be actively "setting" your will in one direction or another, but rather failing or succeeding in doing what your will was already set on doing. I have discussed these and other issues about weakness of will at greater length elsewhere, e.g., in *The Significance of Free Will*, 130–33, 154–57, where I argue that traditional discussions of this topic have overlooked a number of significant distinctions. I am indebted to Chandra Sripada for helpful comments on the original paper related to this topic.

be a *matter of degree* and would be limited by formative circumstances—often severely limited, as in the case of the young man at trial. I thus reject all libertarian theories of the past that make free will an all-or-nothing affair, appealing, for instance, to various forms of agency or causation (noumenal or otherwise) that may somehow escape the clutches of laws of nature and formative influences. I quote here on this topic from the final section of my book *The Significance of Free Will*, written now over twenty years ago.

> Such all-or-nothing views of [libertarian] free will tend to lead in some thinkers to harsh retributive theories of punishment that recognize few circumstances mitigating guilt or responsibility . . . . But if free will is embedded in the natural order, as in the theory of this book, libertarians can avail themselves of . . . everyday intuitions for excusing or mitigating guilt and moral responsibility"—immaturity, childhood abuse, mental disabilities, and the like.[23]

For

> What is needed for a mature libertarianism in the modern age is a recognition of the many ways in which circumstances of birth and upbringing can limit free will and responsibility (for this is one of the prevailing themes of modernity) without yielding to the temptation to think that we are all [always] helpless victims of circumstances.[24]

These themes are deeply reflected in my own personal experiences as well, which were taking place in my life when I was writing that book. My oldest son suffered from schizophrenia, which manifested itself only at age nineteen and until his accidental death at age twenty-seven. It was important to our son that we could communicate to him whenever possible that he was responsible for his behavior and could deal with his affliction by exercising will power and freedom of will, even while recognizing that in moments under stress, we understood that he was not completely in control of his actions. His sense of self-control was essential to his self-image and to sustaining the will power to take his meds, limiting his drinking and smoking, being kind to others, and ultimately dealing as a responsible agent with his affliction. On the night he died, he had earlier called his Mom to proudly announce that though he was out partying, he had only had two beers. It would have been disastrous, I believe, to take a hard line and suggest that because of his affliction, he was not a truly morally responsible agent, at least in many aspects of his life. Martin Voss, an influential neuroscientist at the University of Berlin, who has studied

[23] Kane, *The Significance of Free Will*, 214.
[24] Ibid.

action deficits in schizophrenics, told me at dinner at a conference in Germany some years ago, that in his view this was the appropriate attitude to have taken to our son.

Similarly, on the other side of the moral ledger, if we want to reign in the exploiters and other evildoers of this world, the last thing we should be telling them is that they are not to any degree ultimately responsible for being the way they are.

Indeed, I also make this idea that responsibility is a matter of degree crucial to my defense of libertarian free will. Following the above quotes at the end of *The Significance of Free Will* in this regard, I express agreement with Hans Blumenberg's assertion, in his book *The Legitimacy of the Modern Age*,[25] that "while the quest for autonomy is integral to the modern age . . . complete autonomy is an impossible ideal" and I add the following:

> If we want to be independent sources of activity in the world, we must accept ambivalence, uncertainty, struggle, and conflict within ourselves—all of which are connected to the indeterminacy that is required for free will. The ambivalence, uncertainty, and risk are in turn related to competing images of the good that must inevitably confront those who would be [at least to some degree] ultimate creators of their own ends.[26]

Philosophy, Law, University of Texas at Austin

[25] Hans Blumenberg, *The Legitimacy of the Modern Age* (Cambridge, MA: MIT Press, 1983), 65.
[26] Kane, *Significance of Free Will*, 214–15.