# 1

# *Libertarianism*

*Robert Kane*

COPYRIGHTED MATERIAL

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

*Jalalu'ddin Rumi, thirteenth-century Persian poet and mystic*

## 1 Introduction: An Ancient Problem with Modern Significance

The problem of free will and necessity or determinism of which Jalalu'ddin Rumi spoke has arisen in history whenever humans have reached a higher stage of self-consciousness about how profoundly the world may influence their behavior in ways unknown to them and that they do not control. The rise of doctrines of determinism or necessity in the history of ideas is an indication that this higher stage of self-consciousness has been reached. People have wondered at various times whether their actions might be determined by Fate or by God, by the laws of physics or the laws of logic, by evolution, genes or environment, unconscious motives, upbringing, psychological or social conditioning, or, with the latest scientific threats from the neurosciences, by the activity of the neurons of their brains of which they are not conscious.

There is a core idea running through all these historical doctrines of determinism or necessity, whether they are religious, secular, or scientific,

*Four Views on Free Will*, Second Edition. John Martin Fischer, Robert Kane, Derk Pereboom, and Manuel Vargas.
© 2024 John Wiley & Sons Ltd. Published 2024 by John Wiley & Sons Ltd.

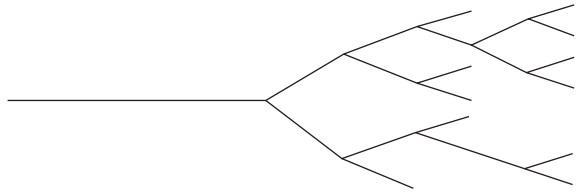

**Figure 1.1** Garden of forking paths.

that shows why many people have felt they are a threat to free will. This core idea may be stated as follows:

> **Determinism**: given the past at any time and the laws governing the universe, there is only one possible future. Whatever happens is therefore inevitable; it cannot but occur, given the past and laws.
>
> **Free will**: by contrast, this implies (i) an open future, with multiple possible paths into the future and that (ii) it is sometimes "up to us" which of these possible paths we will take.

Such a picture of an open future that free will seems to require is often illustrated by an image made famous in a short story by the well-known South American writer Jorge Luis Borges. It is the image of a "garden of forking paths" illustrated in Figure 1.1. At each juncture there are forking paths into the future. If we believe our choices about which of these paths we will take at such times are *free* choices, we must believe both options are "open" to us while we are deliberating. We could choose different paths into the future at various points in our lives; and it would be "up to us" and no one and nothing else which of these paths will be taken.

I believe such a picture of different possible paths into the future, at least at some times in our lives, is essential to our understanding of free will. Such a picture is also important, we might even say, to what it means to be a person and to live a human life. Yet determinism, if true, would seem to threaten this picture. For it implies that there really is, at all times, only one possible path into the future, not many. We may *believe* there are multiple paths available to us. But in reality, if determinism is true, only one of them would be possible.

## 2 Modern Debates and Views

Like Rumi and many other thinkers of the past, I had always believed there was some kind of conflict lurking here that was very deep and could not be easily dismissed by facile arguments. Yet I was also aware that many philosophers and scientists, especially in the modern era, have argued that doctrines of determinism

pose no real threat to free will, or at least to any free will "worth wanting." These thinkers are usually called "compatibilists":

> **Compatibilists**: believe that free will is compatible with determinism, so that we can have all the free will that is possible and worth wanting, even if determinism should be universally true.

Even in a determined world, these compatibilists argue, we would want to distinguish persons who are free from such things as physical restraint, addiction or neurosis, coercion, compulsion, covert control by others, or political oppression from persons who are not free from these things; and we could affirm that these freedoms could exist and would be preferable to their opposites *even in a determined world*. In addition, these modern compatibilists commonly argue that requiring that free actions must be *un*determined would not do anything to enhance our freedom, but would rather reduce our freedom to mere chance or luck or mystery.

In modern debates about free will, compatibilist views of these kinds are opposed by:

> **Incompatibilists**: those who deny that every kind of freedom "worth wanting" is compatible with determinism.

I will be defending such an incompatibilist view in this debate. While many kinds of freedom may be compatible with determinism, as the preceding paragraph suggests, I believe there is one important kind of freedom – traditionally called the "freedom of the will" – that is also worth wanting but that is not compatible with determinism.

Freedom of will of this incompatibilist kind satisfies the two conditions mentioned earlier that seem to be threatened by determinism, that is (i) at least at some points in our lives, we face a genuinely open future, with forking paths into that future, either of which we may choose, and (ii) at these crucial times, it is "up to us" and no one and nothing else, which of these possible paths into the future will be taken. We determine our future at such times and the kinds of persons we will become. Those who believe there is an important kind of freedom of will that we can possess satisfying these conditions which is not compatible with determinism are usually called "libertarians" about free will in contemporary debates (from the Latin *liber* meaning "free").[1]

> **Libertarians about free will** believe there is an important kind of freedom of will we can possess that is *incompatible* with determinism and satisfies the following conditions: (i) at some points in our lives, we face a genuinely open future, with forking paths into that future, either of which we may choose, and (ii) at these

> crucial times, it is "up to us," and no one and nothing else, which of these possible paths into the future will be taken.

I will be defending such a libertarian and incompatibilist view of free will in this debate. Many thinkers believe that a free will of the kind libertarians defend – a free will that is not compatible with determinism – is not even possible or intelligible. It is not a kind of freedom, they argue, we *could* have. This worry has a long history and is related to an ancient dilemma: if free will is not compatible with determinism, it does not seem to be compatible with *indeterminism* either. Arguments have been made since the time of the ancient Stoics that *undetermined* events would occur spontaneously and hence could not be controlled by agents in the way that free and responsible actions would require.

If, for example, a choice occurred by virtue of some undetermined quantum events in one's brain, it would seem to be a fluke or accident rather than a responsible choice. Undetermined events occurring in brains or bodies, it is commonly argued, would not seem to enhance our freedom and control over, and hence responsibility for, actions, but rather to diminish freedom, control, and responsibility. Arguments such as these and many others have led to often-repeated charges throughout history that undetermined choices or actions such as a libertarian free will would require would be "arbitrary," "random," "irrational," "uncontrolled," "mere matters of luck," or "chance," and hence could not be free and responsible actions at all.

In response, libertarians about free will throughout history have often appealed to special and unusual forms of agency or causation to explain undetermined free actions, while their opponents have cried magic or mystery. Indeterminism might provide "causal gaps" in nature, libertarians frequently reasoned, but that was only a negative condition for free will. Some special form of agency or causation was needed that went beyond familiar modes of causation in the natural order to "fill" those causal gaps in nature left by indeterminism. And thus we had historical appeals to "extra factors," such as noumenal selves outside space and time (e.g. Immanuel Kant); immaterial minds (e.g. René Descartes); or uncaused causes, nonevent agent causes, or prime movers unmoved that might account for an otherwise undetermined free will.

A tempting way to think, to be sure. But such traditional ways of thinking have also prompted charges by compatibilists and free will skeptics and many other modern critics of libertarian free will. These critics argue that one cannot make sense of an undetermined free will without appealing to magical or mysterious forms of agency which have no place in the modern scientific picture of the world and of human beings.

Friedrich Nietzsche summed up this prevailing modern skepticism in his inimitable prose when he said that such a traditional notion of freedom of the will that would underwrite an ultimate responsibility (UR) for our actions and

require that one somehow be an undetermined "cause of oneself" was "the best self-contradiction that has been conceived so far" by the human mind (1989: Section 17.8).

I agree that a traditional idea of free will that would require its being incompatible with determinism is likely to appear utterly mysterious and unintelligible in a modern context unless we learn to think about it in new ways, hence my long struggle in attempting to defend and make sense of such an idea of free will without reducing it to mere chance, on the one hand, *or* to mystery, on the other. Yet the struggle seemed worth the effort. For, like many another issue of modernity, the question is whether *something* of this traditional idea of free will in what Nietzsche called "the superlative metaphysical sense" can be retrieved from the dissolving acids of modern science and secular learning. Or would it become, along with other aspects of our self-image, yet another victim of the "disenchantments" of modernity?

Yet I came to realize that any retrieval of this idea of free will that would require its being incompatible with determinism would be no simple matter, if it were possible at all. Such a retrieval would require answering not one question but a whole host of questions. And it would require rethinking the relations of many different and related notions: agency, choice, mind, action, selfhood, will, control, responsibility, power, and many others.[2] I will be addressing many of these questions and topics here, beginning with the following central question in contemporary debates about free will.

## 3 The Compatibility Question, Alternative Possibilities, and Ultimate Responsibility

Why might one believe there is an important kind of free will worth wanting that is not compatible with determinism? The first step in answering this question is recognizing that – as this so-called Compatibility Question is usually formulated in many modern discussions of free will, "Is *freedom* compatible or incompatible with determinism?" – the question is too simple. For, as noted in Section 2, *there are many meanings of "freedom"* (as one would expect from such a much-disputed and debated term) and many of these meanings *are* compatible with determinism. Even in a determined world, as noted, we would want to distinguish persons who are *free* from such things as physical restraint, addiction, coercion, and political oppression from persons not free from these things. And we should acknowledge that these freedoms *are* significant ("worth wanting") – having them would be preferable to their opposites – even in a determined world.

Those of us who are libertarians about free will (who believe in a free will that is incompatible with determinism) should, I contend, concede this point to compatibilists: many freedoms worth wanting are compatible with determinism.

What **libertarians about free will** should insist upon is that *there is at least one kind of freedom that is also worth wanting and is not compatible with determinism*. This further freedom is **freedom *of will***, which I define as: "**the power to be the ultimate source and sustainer to some degree of one's own ends or purposes**."

To understand what this notion of free will amounts to, return to the two features mentioned in Section 2 that have historically led persons to believe that free will is threatened by determinism. We believe we have free will when we view ourselves as agents capable of influencing the world in various ways. Open alternatives seem to lie before us (a "garden of forking paths" in the earlier image). We reason and deliberate among them, and choose. We feel (i) it is "up to us" what we choose and how we act, and this means we "could have chosen or acted otherwise" or that we had "alternative possibilities." This "up-to-us-ness" also suggests that (ii) the ultimate sources of our actions lie to some degree in us and not entirely outside us in factors beyond our control.

Most modern debates about whether such a free will is or is not compatible with determinism have tended to focus on the first of these two requirements, which might be called the:

> **Condition of alternative possibilities (AP)**: free agents must have "alternative possibilities" or "open alternatives" for choice or action, which implies that they "could have chosen or acted otherwise."

But arguments about whether or not this much-discussed condition of alternative possibilities (AP) is compatible with determinism have led to contentious debates in modern philosophy. These debates in turn have tended to stalemate over differing interpretations of what it means to say that agents have alternative possibilities, or that agents "could have done otherwise" than they actually did, or that they had the "power" or "ability" at a given time to act *or* to act otherwise.

I believe these contentious debates about the meaning of such expressions as "could have done otherwise" and the resulting stalemates about the role of this AP condition in modern debates about free will are symptoms of a deeper problem. The deeper problem is that focusing on alternative possibilities alone is too thin a basis on which to rest the case for the incompatibility of free will and determinism. It is not that alternative possibilities and the power to do otherwise are unimportant for free will. Far from it. They are very important and we will return shortly to consider why. It is rather that other considerations must also be brought into the picture in arguing for the incompatibility of free will and determinism if we are to fully understand historical and contemporary debates about free will. The Compatibility Question concerning free will and determinism cannot be resolved by focusing on alternative possibilities or the power to do otherwise alone.

Realizing this, I have argued that one must revisit the long history of debates about free will to see where else to look. When doing so, one finds there is another historical condition fueling incompatibilist intuitions that to my mind is even more important than the alternative possibilities condition. This other condition is related to the second of the two requirements for the "up-to-usness" of freedom of will mentioned, namely that "the ultimate sources of our actions must lie to some degree in us and not entirely outside us in factors beyond our control." I call this further condition:

> **The condition of ultimate responsibility (UR)**: the basic idea is that to be ultimately responsible for an action, an agent must be responsible to some degree for anything that is a sufficient reason (a sufficient condition, cause, or motive) for the action's occurring.

If, for example, a choice were to issue from, and can be sufficiently explained by, an agent's character and motives (together with background conditions) at the time, then to be ultimately responsible for the choice, the agent must be at least in part responsible by virtue of choices or actions voluntarily performed in the past for *having* the character and motives he or she now has. Compare Aristotle's claim that if a man is responsible for wicked acts that flow from his character, he must at some time in the past have been responsible for forming the wicked character from which these acts flow (Aristotle 1985: 67).

This condition of UR accounts for the *ultimate* in the original definition of free will given earlier: "the power of agents to be the ultimate sources and sustainers to some degree of their own ends or purposes."

## 4 Self-forming Actions

Importantly, such a condition of ultimate responsibility (UR) does not require that we could have done otherwise (AP) for *every* act done "of our own free wills." This UR condition thus partially vindicates compatibilists and others who insist that we can be held morally responsible for many acts even when we could not have done otherwise than perform them. But the vindication is only partial. For this UR condition *does* require that we could have done otherwise with respect to *some* acts in our life histories by which we *formed* or *shaped* our present characters, motives, and purposes (i.e. our wills). I call these character and will-forming actions:

> **Self-forming actions (SFAs)**: those acts by which we form and reform our wills (our characters, motives, and purposes) and for which we could have done otherwise, that must occur at some times in our lives, if we are to be ultimately responsible for having the wills we have and hence for being the kinds of persons we become.

To bring out the importance of these self-forming actions (SFAs), consider a familiar line of argument purporting to show that moral responsibility does not require alternative possibilities or the power to do otherwise *at all*, a line of argument illustrated by compatibilist Daniel Dennett's much-discussed example of Martin Luther (1984: 131–133). When finally breaking with the Church at Rome, Luther said, "Here I stand, I can do no other." Suppose, says Dennett, Luther was literally right about himself at that moment. Given his character and motives, he literally could not then have done otherwise. Does this mean he was not morally responsible? Not at all, Dennett says. In saying, "I can do no other," Luther was not disowning responsibility for his act, but taking full responsibility for it. Thus, compatibilist Dennett concludes, "could have done otherwise (AP) is *not* required for free will in a sense demanded by moral responsibility."

In response, I would argue that incompatibilists about free will may, and indeed should, grant that Luther could have been responsible for this act, *even ultimately responsible in the sense of UR*, though he could not have done otherwise then and there, even if his act were determined by his existing will at that moment. But this would be so, incompatibilists should argue, to the extent that Luther was responsible for his present *motives and character* (his will) by virtue of some earlier struggles and self-forming choices (SFAs) that *brought him to this point where he could do no other*.

> **Acting of "our *own* free will"**: often we act from a will already formed, but it is "our *own* free will" by virtue of the fact that we formed it by other choices or actions in the past (*self-forming choices or actions* or SFAs) for which we could have done otherwise. If this were not so, *there is nothing we could have ever done differently in our entire lifetimes to make ourselves different than we are* – a consequence, I believe, that is incompatible with our being (at least to some degree) ultimately responsible for being the way we are.

## 5 Freedom of Action and Freedom of Will: AP and UR

Focusing on this condition of ultimate responsibility (UR) tells us something else of importance about the traditional problem of free will. It tells us why it is a problem about the freedom *of the will* and not just about the freedom *of action* and why these freedoms must be distinguished if the Compatibility Question and other questions about free will are to be adequately addressed.

There has been a tendency in the modern era of philosophy, beginning with Thomas Hobbes and John Locke in the 17th century, and coming to fruition in the 20th century, to reduce the problem of free will to a problem of free action. I believe such a reduction oversimplifies the problem.

*Free will is not just about free action*, though it involves free action. *Free will 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. Were we ultimately responsible to some degree for having the wills (characters, motives, and purposes) we do have, or can the sources of our wills be completely traced back to something over which we had no control, such as Fate or the decrees of God, heredity and environment, social conditioning, or hidden controllers, and so on? Therein, I believe, lies the core of the traditional problem of "free will."

John Locke (1690/1975, Book II, Chapter xxi: 134) famously said in the 17th century that the so-called problem of free will which had so exercised medieval and earlier philosophers was really a problem about *free agency*, or the freedom *of the agent*, and not about the freedom *of the will*. Like other thinkers of the modern era, Locke was skeptical of medieval references to the "will" in general which often made it out to be a mysterious inner homunculus or power capable of influencing actions and events in ways not countenanced, Locke believed, by the new emerging sciences of his day.

Many moderns down to the present time have followed Locke in this skepticism. They argue, as did Ludwig Wittgenstein (1953) and Gilbert Ryle (1949) in the mid-20th century, that references to the *will* and *acts of will* were outdated remnants of pre-modern modes of thought and should go the way of witches and phlogiston. Even some modern libertarians and incompatibilists about free will have joined compatibilists in arguing that the historical "problem of free will" is really about the freedom of agency and hence about the freedom of action and not about the freedom of the will.

I believe this is a mistake. It is *not* wrong, to be sure, to say as Locke does that the traditional problem of free will is really a problem about *free agency*. But it *is* wrong to say it is not therefore *also* about the freedom of the will. For, as described in these paragraphs, *freedom of will is an important aspect of free agency*; and moreover *free will is that particular aspect of free agency that has been the subject of historical debates about whether it is or is not compatible with determinism*.

For if the case for the incompatibility of free will and determinism cannot be made by appealing to the condition of alternative possibilities (AP) alone, the case can be made if UR is added. I have thus argued that UR should be moved to center stage in free will debates. To be ultimately responsible for an action in the sense required by UR, an agent must be responsible to some degree for anything that is a sufficient reason (cause or motive) for the action's occurring. And this implies, as noted, that if a choice or action can be sufficiently explained by an agent's present character, motives, and purposes, then to be ultimately responsible for the choice or action, the agent must be at least in part responsible by virtue of choices or actions voluntarily performed in the past for *having* the character, motives, and purposes he or she now has. But this being the case, an

impossible infinite regress of past choices or actions would be required unless *some* choices or actions in the agent's life history (self-forming actions, or SFAs) did *not* have *sufficient* causes or motives (and hence were not determined).[3]

Yet if one could arrive at the incompatibility of free will and determinism from this condition of ultimate responsibility (UR) alone in this manner, one might wonder whether appeals to alternative possibilities (AP) are needed at all for free will. Some recent philosophers who are impressed by arguments of the above kind for incompatibilism (they are sometimes called "source incompatibilists") have suggested that appeals to alternative possibilities and the power to do otherwise are not needed at all for free will.

I believe this is also a mistake. *Both conditions – UR and AP – are needed for free will*. But the reasons why both conditions are needed are more subtle than is generally realized; and understanding them requires further steps in rethinking the Compatibility Question.

## 6 Plurality Conditions and Plural Voluntary Control

The first of these further steps concerns what I call "plurality conditions" for free will. When we wonder about whether the wills of agents are free, it is not merely whether they could have done otherwise that concerns us, *even if the doing otherwise is undetermined*. What interests us is whether agents could have done otherwise *voluntarily* (or *willingly*), *intentionally* and *rationally*. Or to put it more generally, we are interested in whether agents could have acted voluntarily, intentionally, and rationally *in more than one way*, rather than in only one way, and in other ways merely by accident or mistake, unintentionally, involuntarily, or irrationally. I call such conditions:

> **Plurality conditions for free will** (Kane 1996: 107–111): the power of agents to act voluntarily, intentionally, and rationally *in more than one way*, rather than in only one way, and in other ways merely by accident or mistake, unintentionally, involuntarily, inadvertently, or irrationally.

Such conditions seem to be deeply embedded in our intuitions about free choice and action. Most of us naturally assume that freedom and responsibility would be deficient if it were always the case that we could *only* do otherwise by *accident* or *mistake*, *unintentionally*, *involuntarily*, or *irrationally*.

To illustrate, imagine a world in which there is a considerable amount of genuine indeterminism or chance in human affairs as well as in nature. In this world, people set out to do things – kill prime ministers, press buttons on machines, punch computer keys, hit targets, etc. – usually succeeding, but sometimes failing by mistake or accident. Suppose an assassin, who usually

hits his targets, is aiming to kill a prime minister from a distance with a high-powered rifle, when some undetermined events in his nervous system lead to a wavering of his arm and he misses his target. Or suppose I approach a coffee machine meaning to push the button for black coffee when, due to an undetermined brain cross, I accidentally press the button for coffee with cream.

Now imagine further that *all* actions in this world in the lifetimes of agents, whether the agents succeed in their purposes or not, are such that their reasons, motives, and purpose*s* for wanting and trying to act as they do are always preset or settled by prior circumstances of heredity, environment, social conditioning, and other formative circumstances. Whether the assassin misses the prime minister or not, his *intention* to kill is already *settled* prior to his attempt, by his past formative circumstances. Whether I succeed in pressing the button for coffee without cream, my wanting to do so because of my dislike of cream is already *settled* by my formative circumstances. And so it is, we are to assume, for all persons and all their actions in this imagined world.

I would argue that persons in such a world lack free *will*, even though it may often be the case that they have (i) *alternative possibilities* and that their actions are (ii) *undetermined*. For they can sometimes do otherwise than they do in a manner that is undetermined, but only inadvertently or unintentionally, by mistake or accident, as in the case of the assassin or my pressing the wrong button on the coffee machine – and this is a limited kind of freedom at best. What they cannot do is *will* otherwise than they do. Their reasons, motives, and purposes have been already "set one way" before and when they act, so that if they act otherwise, it will *not* be "*in accordance with their wills*," but rather by chance or accident.

What this shows is that when we wonder about whether the *wills* of agents are free, it is not only whether they could have done otherwise that concerns us, even if their doing otherwise is undetermined. What interests us is whether they could have done otherwise *voluntarily* (in accordance with their wills), *intentionally* (knowingly rather than inadvertently and on purpose rather than accidentally), and *rationally* (having reasons for so acting and acting for those reasons). Or to put it more generally, we are interested in whether they could at some times have acted voluntarily, intentionally, and rationally *in more than one way*, rather than in only one way, and in other ways merely by accident or mistake, unintentionally, inadvertently, or irrationally.

We thus arrive at an answer to the question of why these "plurality conditions" are so deeply embedded in our intuitions about free choice and action. We naturally assume that freedom and responsibility would be deficient if it were always the case that we could *only* do otherwise by accident or mistake, unintentionally, involuntarily, or irrationally, in short, *unwillingly*. To have freedom of will, we must not only be able to do otherwise: we must be able to do otherwise *willingly* or *at will*. If free *will* involves more than alternative possibilities and indeterminism, these plurality conditions appear to be among the significant additional requirements.

Reflecting on these plurality conditions tells us something else of importance about free will. For, satisfying such plurality conditions implies that agents must be able to exercise a certain kind of control over some of their actions that I refer to as:

> **Plural voluntary control (PVC)**: agents have PVC over a set of options (e.g. choices or actions) when (i) they are able to bring about *either* of the options *voluntarily* (without being coerced or compelled or otherwise controlled by other agents or mechanisms), *intentionally* (knowingly and on purpose, rather than merely by accident or mistake) and *rationally* (for *reasons* that they then and there wish to act upon) and (ii) *whichever* option they do bring about by exercising such PVC will have been brought about by them voluntarily, intentionally, and rationally in these senses.

These conditions can be summed up by saying, as we sometimes do, that the agents can act or choose either way "at will" or, alternatively, that it is "up to them" which way they will choose or act when they choose or act.

## 7 Will-setting and Self-formation

Focusing in this way on plurality conditions and plural voluntary control (or PVC) also leads to a further important and often-neglected topic in free will debates that I call "will-setting" (Kane 1996: 113–115). In the imagined scenario in Section 6, all of the motives and purposes of agents in every situation are already "preset" or "set one way" before they act. The assassin's desires and purposes are set on killing the prime minister, not on missing or killing an aide. My desires and purposes are set on pressing the button for black coffee and not any other button. In such cases, where the motives and purposes of agents are already set one way before they act, we may say:

> Their actions are **will-settled**, meaning that the wills of agents, their motives and purposes, are already set one way on doing something *before* they act.

By contrast:

> Actions are **will-setting** when the wills of agents, their motives and purposes, are *not* already preset or set one way *before* they act. Rather the agents set their wills one way or another *in the performance of the actions themselves*.

Choices or decisions, which are self-forming actions (SFAs), in the sense defined here, are will-setting in this sense. The agents' wills are not already set one way before they choose, but they set their wills, one way or the other,

voluntarily, intentionally, and rationally, *in the act of choosing itself*. Such self-forming actions would thus satisfy the plurality conditions.

The imagined world in which all the motives and purposes of agents are already set one way whenever they act thus provides a clue to the deep connection between will-setting, ultimate responsibility (UR), free will, and the plurality conditions. If we are to be to some degree ultimate determiners of our own wills, as UR requires, some actions in our lifetimes (self-forming actions, or SFAs) must be will-setting in the above sense and hence must satisfy the plurality conditions. But these self-forming actions will then satisfy the condition of alternative possibilities (AP) as well. For if one can do or do otherwise, voluntarily, intentionally, and rationally either way, it follows that one can do or do otherwise. One has alternative possibilities. AP would therefore be necessary for free will after all, *at least sometimes in our lives when we engage in self-formation*.

## 8 The Compatibility Question Revisited: Free Will and Moral Responsibility

Focusing on both ultimate responsibility (UR) and alternative possibilities (AP) when discussing the Compatibility Question, rather than merely on AP alone, has another significant consequence. It shows why issues about free will have been so deeply entangled throughout history with issues about *moral responsibility* for actions. This entanglement is no accident. It has to do with the very meaning of freedom *of will* (which involves both UR and AP). Reflecting on this entanglement of free will and moral responsibility leads to further arguments relating to whether freedom of will is or is not compatible with determinism – arguments having to do with our ordinary practices of holding persons responsible for their actions in everyday moral and legal contexts.

Many contemporary compatibilists and other philosophers have been influenced on these topics by a seminal 1962 article by British philosopher P.F. Strawson. In this influential article, entitled "Freedom and Resentment," Strawson focused on our ordinary practices of holding persons morally responsible and on what he called the:

> **Reactive attitudes**: attitudes toward persons usually associated with ordinary practices of holding persons morally responsible, including attitudes such as blame, resentment, indignation, guilt, moral approval, and moral praise.

Strawson argued that our ordinary practices of holding people responsible, including these reactive attitudes, were basic to our human form of life and could be wholly "insulated" from traditional abstract philosophical and scientific

concerns about free will and determinism. To believe, he argued, that our ordinary practices of holding persons responsible in everyday life and the reactive attitudes related to them would have to be qualified in some ways – or even possibly abandoned – if we found that all their actions were determined by prior causes was to "overintellectualize" the issues.

This "insulation thesis" (as it has sometimes been called) is a controversial feature of Strawson's article, and it has had numerous proponents and critics. Interestingly, one of the most prominent of the critics was Strawson's son, Galen Strawson, who in his 1986 book *Freedom and Belief* took issue with his father's contention in "Freedom and Resentment" that ordinary practices of blaming and other reactive attitudes *could* be entirely "insulated" from metaphysical concerns about determinism. Against this contention, Galen Strawson argued that "the roots of the incompatibilist intuition" (that free will is incompatible with determinism) "lie deep" in our ordinary practices and in the reactive attitudes associated with those practices. These ordinary practices and the reactive attitudes associated with them, he argued, "enshrine the incompatibilist intuition," rather than being "insulated" from it (Strawson 1986: 89).

I agree with Galen Strawson on this issue, though my reasons are not all the same as his. Like him, I believe our ordinary practices of holding persons morally responsible and related questions about blameworthiness and the reactive attitudes cannot be *entirely* insulated from philosophical worries about free will and determinism that have engaged philosophers for centuries. This genie cannot be kept in the bottle, annoying as he may be. There are a number of ways to show this that I will explore in Section 9.

## 9 Fair Opportunity to Avoid Wrongdoing: Hart and Others

The first way focuses on ordinary practices of ascribing responsibility, culpability, and blame in courts of law and other legal contexts. A widely cited condition among legal theorists for such ascriptions was stated by the influential British legal theorist H.L.A. Hart (1968). It may be called:

> **The fair opportunity condition**: a necessary condition for ascribing responsibility and culpability to agents in legal contexts according to which the agents must have had a "fair opportunity to avoid wrongdoing," or more generally, a "fair opportunity to have done otherwise" than they have done.

In an important article, David Brink and Dana Nelkin (2013) argue persuasively that this "fair opportunity" criterion of Hart's is not only crucial for understanding legal and criminal responsibility but also crucial for understanding moral responsibility in general, in an accountability sense that would justify

blame, sanction, and punishment. Hart's fair opportunity condition, they argue, is thus a crucial part of the "architecture" of ordinary practices of ascribing moral as well as legal responsibility (Brink and Nelkin 2013: 284).

If this is the case, as I believe it is, it has implications concerning whether our ordinary practices of ascribing responsibility can *be* insulated from traditional philosophical debates about free will and determinism. For, appeals to Hart's fair opportunity criterion for assigning responsibility in ordinary legal and moral practices lands one squarely in the center of traditional philosophical debates that have concerned incompatibilists about whether causal determinism rules out the freedom to do otherwise and whether and to what degree the freedom to do otherwise is required for moral responsibility.

To show this, it is instructive to consider the following: if causal determinism were 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* in the following sense:

> **Causal impossibility**: an event E's occurring at a time t is *causally impossible* just in case the following is true: "If the past prior to t is as it is in the actual world and the laws of nature are as they are in the actual world, then E cannot possibly occur at t."

If causal determinism was true, *anything you might have done differently in the course of your life* to make yourself different than you are at any time – your character, your motives, your dispositions, your intentions, the quality of your will – would have been causally impossible in this sense. For, as we have seen in Section 1, causal determinism implies that given the past and the laws governing the universe at any time, there is only one causally possible future. And if an agent's avoiding wrongdoing was causally impossible in this sense, it would certainly appear that the agent lacked a "fair opportunity" to avoid doing it.

An important qualification, however, must immediately be added here – a qualification that is crucial for understanding not only Hart's fair opportunity criterion but also ascriptions of responsibility in ordinary moral contexts generally. The qualification is that the causal impossibility of avoiding wrongdoing *in certain particular circumstances* will not always imply that agents are excused from moral responsibility *in those circumstances*. It implies this only in certain conditions.

For example, if 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 the driver of responsibility. For one must also ask whether the driver 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

have avoided the accident, such as his prior decisions to drink and then drive. The causal impossibility of avoiding doing something *now* (e.g. avoiding the accident when it occurred) will not excuse an agent of responsibility, if some of the crucial circumstances that made it now causally impossible to avoid doing it were the results of actions or omissions by the agent in the past, *which the agent had a "fair opportunity to avoid" when they occurred*. And this last phrase is crucial. 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 doing when *they* occurred.

## 10 Reactive Attitudes, Criminal Trials, and Transference of Responsibility

Another significant way of highlighting problems with the thesis that issues about responsibility and the reactive attitudes can be insulated from philosophical concerns about determinism is discussed by a number of other writers. It is discussed, for example, by Shaun Nichols in his book *Bound: Essays on Free Will and Responsibility* (2015). Nichols's book focuses on the implications of new research in empirical psychology and experimental philosophy for traditional philosophical debates about free will and moral responsibility. The passages of interest here in his book are where Nichols discusses Galen Strawson's claim, mentioned in Section 8, that our ordinary practices of holding responsible, and the reactive attitudes related to them, "enshrine" incompatibilist intuitions about freedom and responsibility rather than being entirely "insulated" from such intuitions.

In discussing what he takes to be important arguments supporting this claim of Galen Strawson's, Nichols introduces two examples that play a pivotal role in his discussion. One of these examples is from Gary Watson's (1987) well-known and much-discussed account of the ruthless murderer, Robert Harris, on death row in California for multiple murders. The other example Nichols (2015) considers in his book is taken from my own writing about the trial of a young man who assaulted and raped a teenage girl. The examples have similar import. But I will focus on my example because it brings out some key points that Watson does not emphasize. As Nichols points out, my example is roughly based on experience, triggered by the trial of the young man accused of the assault and rape.

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 his victim and who 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 such factors (not entirely unlike Watson's 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" entirely to others and wondered whether some residual responsibility and blame might not belong to the young man himself. My question became: Was his behavior *all* a matter of bad parenting, neglect, and abuse, and like factors, or did he have any crucial role to play in choosing it?

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 the young man 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 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 the young man to have resisted the influences of his genetics and upbringing* and to have acted differently at some points in his life to make himself different than he now is. And if determinism were true, acting differently than he actually did at *any* time in his lifetime would have been *causally impossible*.

One might argue here 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 not in any way influenced by listening to the history of the young man (as I am sure there were), then I would not want to see those persons 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."

## 11 Transference of Responsibility and Compatibility Questions

It is worth reflecting further on this interesting notion of "transference of responsibility." We are inclined to do this to some degree to other persons who may have influenced agents to be the way they are and act as they do, to the extent we believe that the influences of these others *were difficult for the agents to have resisted when they occurred*. In such cases, we are inclined to "transfer" at least some of the responsibility and blame to those others who so strongly influenced the agents (parents, caregivers, role models, abusers, and so on) and to mitigate the responsibility of the agents accordingly. In some extreme cases, such as Watson's Robert Harris or the young man of my example, we might possibly conclude that the influences were so strong that resisting them to any degree was not causally possible. We could be wrong about this. Such judgments are fallible and should be made with great caution. But the crucial point is that *such judgments are relevant to our ordinary practices of holding agents responsible and blameworthy*, including who should be held responsible and to what degree.

Moreover, as noted, such judgments seem to depend on whether or not it was ever causally possible for the agents to have resisted the influences of their environment and upbringing and to have acted differently at some points in their lives to make themselves different than they are. And if determinism were true, this would never have been causally possible.

There are further interesting and relevant implications of this notion of transference of responsibility for free will debates. Many compatibilists – Daniel Dennett (1984) being a prominent example – are willing to concede that this transference of responsibility and related reactive attitudes, such as blame and resentment, to other persons who may have abused or otherwise exerted powerful influences over agents is indeed a normal feature of our ordinary practices of holding persons responsible. But these compatibilists go on to argue, as does Dennett, that such transference is only reasonable if the responsibility, blame, resentment, etc., is transferred to some other *persons*. If the influences on the agent's behavior and quality of will are due to natural causes alone and no other persons can be implicated, these compatibilists say, then it is not reasonable to transfer responsibility and other reactive attitudes, such as blame and resentment, to nature. For, as Dennett succinctly puts it, "nature is not a person."

This is true enough so far as it goes and should be admitted by incompatibilists. It is an important fact about this transference phenomenon that transference of

moral and legal responsibility must be to other *persons* and cannot be to natural causes alone. Indeed, this fact is related to something important about the reactive attitudes, such as resentment, indignation, and blame, in general, namely that they are appropriately directed only at beings who are themselves capable of *responsible agency*. But rather than providing a decisive argument for compatibilism, further reflection on this significant fact can take us in an opposing direction.

To see why, return for a moment to the young man on trial. Since he seems to have acted voluntarily and intentionally from a perverse and vicious will, our reactive attitudes of blame and resentment were initially focused entirely on him. But when we heard more about his past, we wondered whether some of the blame at least should be transferred to the sexually abusive father and others who may have enabled the father. But now suppose we learn that the abusive father was as he was because *he* was sexually abused by his father and so on back indefinitely. Perhaps it was all in some bad genes. Or suppose the young man's having the perverse will from which he acted was all a matter of determining genetic mutations in his fetal development for which we cannot blame the father *or* any other persons.

Suppose further that, as evidence unveiled in the courtroom made clear, the young man's acting from his perverse will in this incident satisfied familiar compatibilist conditions for free and responsible agency: he acted voluntarily, without being coerced or forced by others; he acted intentionally, knowing exactly what he was doing and doing it on purpose; he was reasons-responsive in the sense that he was calculating and would have altered his behavior appropriately, if his beliefs, desires, and circumstances had been different in various ways. Nor was he acting compulsively, as in Harry Frankfurt's (1971) description of the unwilling addict, who wanted to resist the desire to take a drug, but could not resist taking it anyway. On the contrary, this young man had, in Frankfurt's terms, "the will he wanted to have" and was "wholehearted" and not ambivalent in his commitment to act in accordance with the will he had, perverse though it might be.

It was indeed the evidence of all this, coming to light in the proceedings, that led most of those present in the courtroom to our initial attitudes of resentment and moral anger toward the young man. Most of us, to be sure, transferred these reactive attitudes to some degree to others when we learned more about the sordid details of his upbringing. But we are now imagining a different situation. We are now supposing that the young man's having the perverse will from which he acted was a matter of determining genetic mutations in his fetal development for which we cannot blame an abusive father or any other persons. So we cannot blame any other *persons* for the young man's having the perverse will he does have. Yet it seems we can't blame *nature* either, which is not a person. And this prompts the following reflections.

Imagine two young men, possessing exactly the same wills and motives as this young man, and satisfying all the same compatibilist criteria for responsibility (uncoerced, reasons-responsive, etc.). Yet one of these young men was determined to be so by the actions of other persons, like the abusive father; the other young man was determined to be so by natural, impersonal causes alone (such as mutations in his genetic development). The first young man is not fully responsible, we might say, to the degree that the actions of other persons made him this way. The responsibility transfers at least in part to those others. But then should we say the second young man *is* fully responsible because no other persons, but only natural causes, made him the way he is – in other words, because we cannot find anyone *else* to blame?

Such reasoning seems not only perverse but also completely unfair to the second young man. For the question that begs to be answered *in both cases* is whether it was ever *causally possible* for either young man to have resisted the circumstances that influenced him, and to have made himself different than he turned out to be, whether those influencing circumstances were the result of the actions of other agents or the results of natural causes alone. And if these circumstances were *determining* either way, it would not have been causally possible for either young man to resist them. Neither of the young men would have had a "fair opportunity" to have turned out otherwise.

## 12 Two Dimensions of Responsibility

Reflections such as these suggest that if full justice is to be done to our understanding of moral and legal responsibility and to our practices of holding persons responsible, two dimensions of responsibility must be distinguished. Both dimensions, I would argue, are necessary for a fully adequate account of these practices, and neither dimension 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* and *intentionally*.

The second dimension of responsibility is another matter. It is not responsibility for expressing in action the true quality of will one has or the real self one is, but responsibility for *forming* and thus *having* the quality of will one *expresses* and thus *being* the kind of self one *is*. The distinction put succinctly is between:

> **First dimension of responsibility**: responsibility for *expressing* in action the will one *has*.
>
> **Second dimension of responsibility**: responsibility for *forming* and thus *having* the will one *expresses* in action.

To be responsible in this second dimension it must be that at least some time in one's life, when one acts responsibly and hence voluntarily and intentionally

in the first dimension, it was also possible for one 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 one's will and the self that one was 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, one's *already existing will*. Some choices or actions in one's life must be, as explained in Section 7, *will-setting*, and not already *will-settled*.

## 13 Compatibilist Responses I: Conditional Analyses

Compatibilists are not without further responses to the preceding arguments concerning free will and moral responsibility. Historically, compatibilists have in fact pursued a number of different strategies we have not yet discussed to show that freedom and moral responsibility are compatible with determinism The most common strategy attempting to show this employed by compatibilists in the modern era – from Hobbes and Locke in the 17th century to David Hume and John Stuart Mill in the 18th and 19th centuries, and well into the 20th century, is often called the "classical compatibilist strategy." It involves conceding that moral responsibility *does* require the freedom or power to do otherwise *in some sense*. But then it involves defending *conditional* or *hypothetical* interpretations of the freedom or power to do otherwise, according to which that power turns out to be compatible with determinism.

According to such classical compatibilist strategies, what we mean when we say that agents were "free or had the power to do otherwise," or "could have done otherwise," is that "*they would or might have done otherwise*, *if* the past (or the laws of nature) had been different in some way."

If, for example, persons had had different beliefs or desires, or had reasoned or chosen differently, or were in different circumstances, they would or might have acted differently. And saying persons would or might have acted differently, if the past or laws had been different in some way, these compatibilists then argue, is consistent with saying that the agents acting as they did was determined, given the past and the laws as they actually were.

I believe this standard compatibilist strategy is deeply flawed. Immanuel Kant, as is well-known, called it a "wretched subterfuge" and William James a "quagmire of evasion"; and I think they were right. A number of cogent objections have been made against such conditional interpretations of the freedom to do otherwise since the mid-20th century; and even many compatibilists today express doubts about such analyses. I believe they do so with good reason.

For it may be true that persons would or might have done otherwise, if the past or the laws had been different in some way. But the difficulty is that the actual past when they do act was not different in some way; it was as it was.

Likewise, the actual laws were not different; they were as they were. *Our freedom and responsibility must be exercised in the world that actually is*, *not in some hypothetical or merely possible world that might have been, but never actually was*. And if determinism is true of this actual world in which we live and act, then acting otherwise than we do *in the circumstances we actually find ourselves in* would always be causally impossible. It is not excusing to be told that persons would or might have acted otherwise in some merely hypothetical or possible worlds that never actually existed, if their acting otherwise in the actual world in which they do live and act was causally impossible.

Imagine a 16th-century incarnation of a modern classical compatibilist who took it upon himself to correct Luther when Luther said, "Here I stand. I can do no other." "You were mistaken, sir," this classical compatibilist might have said to Luther. "For all we mean when we say, 'I can now do other' is simply that in some possible worlds in which the past or the laws were different in some ways from the actual past and laws – in which, for example, you had had different beliefs or reasons or purposes – you would or might have done otherwise. And this may well have been true of you, sir" – this classical compatibilist might have said to Luther – "at the time you said, 'Here I stand. I can do no other.' So, you see, you were mistaken at that time to say that you 'could not have done otherwise' at that moment."

Luther would likely have replied, "Get thee gone, sophist! What I meant when I said, 'I can do no other' is that in the *actual world where I found myself at that moment* with all the beliefs, reasons, and purposes I had actually acquired in my long difficult journey to that point, my doing otherwise would have been impossible. What is it to me that I would or might have done otherwise in some merely possible world that did not actually exist at the time. Moreover, it was to some important degree the result *of my own past choices and actions* that the actual world that did exist and in which I acted at that moment was one in which I could not then have done other. For I had brought myself to that point in great part by my own past actions and choices. What you are claiming, therefore, not only distorts what I was saying. It devalues and insults my own contribution to making that actual world in which I was acting at that moment the kind of world in which I could not then have voluntarily and rationally done otherwise."

## 14 Compatibilist Responses II: Frankfurt-style Examples

Most traditional arguments for compatibilism, like the classical compatibilist one considered in Section 13, have conceded that the power to do otherwise *is* required for moral responsibility, but have argued that this power, properly understood, is compatible with determinism. Yet a different and more radical

strategy has become especially popular in contemporary philosophy among compatibilists. It is to argue more directly that the power to do otherwise is not required at all for moral responsibility. The most widely discussed and sophisticated versions of this compatibilist strategy in recent philosophy involve appealing to so-called Frankfurt-style examples (FSEs), named for Harry Frankfurt who formulated the first of these influential examples in 1969.

Frankfurt's aim in formulating the first of these examples was to refute a principle he called:

> **The principle of alternative possibilities (PAP)**: agents are morally responsible for their actions, only if they could have avoided performing them or could have done otherwise when they performed them.

To refute this principle, Frankfurt imagined the following scenario:

> A controller, Black, has direct control over the brain of another man, Jones, and wants to allow Jones to do only what Black wants him to do. Black prefers, however, to allow Jones to act on his own whenever possible and so will only intervene if Jones is about to do something that he, Black, does not want.

Given this scenario, Frankfurt asks us to consider situations in which Jones is about to do what the controller Black wants, so that Black does not intervene. In such situations, Frankfurt argues, Jones could be morally responsible for acting as he does, since he would have acted on his own, from his own motives and for his own reasons, and nothing and no one (including Black) would have interfered with or prevented him from doing what he chose to do.

Yet Jones in such situations, Frankfurt argues, could not have done otherwise, for if he had given any indication of doing otherwise, Black would have prevented him from doing so. Thus, this principle of alternative possibilities (PAP), Frankfurt concluded, is false: it is *not* true that agents can be morally responsible for what they have done, *only* if they could have done otherwise. For such Frankfurt-controlled agents could be responsible for what they have done, if they did it on their own and the controller did not intervene, even though the agents could not have done otherwise, because the controller would have intervened and not let them do otherwise.

Now the first thing to be said about this line of argument is that it should not surprise us at this point that this PAP of Frankfurt's is false. For we have already seen from the discussion of the Luther example and other examples that agents can be morally responsible for actions that flow from their wills at the time they acted and such that they could not have done otherwise at that time. One can be responsible for "will-settled" actions like Luther's "Here I stand," even if one could not have done otherwise than perform them when

they were performed. In other words, we don't need to appeal to unusual examples involving Frankfurt-style controllers to establish that Frankfurt's PAP is false.

But not all of our actions in our lifetimes could be determined or already will-settled in this way when we act, if we are ever to be responsible for our wills being set the way they are when we act. For this to be the case, we would have to, at some times in our lives, be capable of not merely will-settled but also will-setting or self-forming choices or actions that were not determined by our existing wills when we performed them and were such that we could have willingly done otherwise when we performed them. In other words, some, even many, morally responsible actions in the course of our lives may be such that, at the time we performed them, we could not then and there have willingly done otherwise, like Luther's act. But it does not follow that all our morally responsible actions could be like this, if we are ever to be morally responsible to any degree for the state or quality of our wills.

So such examples do show that Frankfurt's PAP is false: it is not true that agents can be morally responsible for their actions, only if they could have avoided performing them, or could have done otherwise, when they performed them. Frankfurt-style examples show this as well and so he and others are right in saying that these examples show that Frankfurt's PAP is false. But such Frankfurt-style examples fail to show the falsity of a more complex principle required for free will that might be stated as follows and that may be called:

> **Will-setting**: agents are ultimately responsible for having the wills (characters, motives, and purposes) they express in action, only if sometimes in their lives they willingly (voluntarily and intentionally) perform certain ("will-setting" or "self-forming") actions (SFAs) that it was causally possible at the time for them to have willingly avoided performing.

These results have more general implications for Frankfurt-style examples. For it can be shown that if all actions were under the control of Frankfurt controllers or mechanisms as in such examples, there could be no such self-forming choices or actions (SFAs) and hence no will-setting of the kind required for agents to be responsible for having the quality of wills they do have.

This is the case because, in all Frankfurt-style examples, including all the more sophisticated versions proposed since Frankfurt's original one, the one thing the Frankfurt controllers can never allow is the following: Frankfurt controllers can never allow the agents an opportunity to bring to completion "will-setting" or "self-forming" actions such that the agents are able to willingly perform the actions and are also able to willingly do otherwise. The controllers must always intervene in some way to prevent such will-setting actions from occurring so that *they* themselves, *the controllers and not the*

*agents*, can *ensure* the agents always do what the controllers want them to do. For the essence of a will-setting or self-forming action is that *the agent, and no one or nothing else*, can determine how such a will-setting action will turn out when it is performed.

## 15 The Intelligibility Question

We arrive finally at the most common and powerful objection made against libertarian views of free will that require its being incompatible with determinism. This objection has been made in various forms throughout history and continues to be commonly made and widely accepted in the present age. It is the objection that a libertarian and incompatibilist free will requiring UR is not even *intelligible* or *possible*. It is *not* something we *could* have anyway.

The culprit here is not determinism, but indeterminism. For, as noted in Section 1, this objection is related to an ancient dilemma: if free will is not compatible with determinism, it does not seem to be compatible with *indeterminism* either. Arguments have been made since ancient times, as noted in Section 3, that undetermined events would occur spontaneously and hence could not be controlled by agents in the way that free and responsible actions would require. If a choice occurred by virtue of some undetermined events in one's brain, it would seem a fluke or accident rather than a responsible choice. Undetermined events occurring in brains or bodies, it is commonly argued, would not seem to enhance freedom and control over, or responsibility for, actions but rather to diminish freedom, control, and responsibility.

In response to such arguments, as also noted, libertarians about free will have often appealed throughout history to special and unusual forms of agency or causation to explain undetermined free actions. Libertarians have appealed to noumenal selves outside space and time, to immaterial minds, uncaused causes, nonevent agent causes, and the like, to account for an otherwise undetermined free will. And their critics have responded in turn that these appeals reinforce the critics' view that one cannot make sense of an undetermined free will without appealing to unusual forms of agency that have no place in the modern scientific picture of the world and of human beings.

These debates thus lead us to another central question concerning libertarian free will that must now be considered. It may be called:

> **The Intelligibility Question**: Is a libertarian free will requiring UR even *intelligible* or *possible*? Can one make sense of such a free will requiring indeterminism without reducing free will to mere *chance*, on the one hand, *or* to *mystery*, on the other, and can such a free will be reconciled with modern scientific views of the cosmos and of human beings?

## 16 Indeterminism: Empirical and Philosophical Questions

In approaching these questions, let us first be clear that it is an empirical and scientific question whether any indeterminism *is* there in nature in ways appropriate for free will – in the brain, for example. No purely philosophical theory alone can settle the matter. As the Epicurean philosophers said centuries ago, if the atoms don't "swerve" in undetermined ways, and in the right places, there would be no room in nature for free will. Christoph Koch is a distinguished neuroscientist and a tough-minded one at that. He argues that "there is no evidence that any components of the nervous system – a warm and wet tissue strongly coupled to its environment – display quantum entanglement" (2009: 40). But Koch goes on to say that "what cannot be ruled out," however, "is that tiny quantum fluctuations deep in the brain are amplified by deterministic chaos" so that they might have nonnegligible nondetermined effects on neural processing and thereby affect human decision-making (2009: 40). Koch does not endorse this idea, but says it cannot be ruled out, given what is currently known about the brain. And such a role for indeterminism is all that would be needed for the view to be presented here.

In the most recent edition of *The Oxford Handbook of Free Will* (2011), Robert Bishop agrees with Koch and cites a number of other neuroscientists and philosophers who have made similar suggestions. If minute quantum indeterminacies occurred at the intraneural or synaptic levels of the brain, affecting the timing of firing of individual neurons, Bishop argues, these indeterminacies, however minute, could be amplified, due to sensitivity to initial conditions, so that they had nonnegligible effects on neural processing in the form of neural noise. Bishop goes on to point out that one need not even appeal only to chaos to get these effects. For, as he notes: "the exquisite sensitivity needed for . . . the amplification of quantum effects is a general feature of nonlinear dynamics and is present whenever nonlinear effects are likely to make significant contributions to the dynamics of a system" (Bishop 2011: 91). And it is generally agreed, Bishop adds, that nonlinear dynamics is pervasive in the functioning of human brains.

A growing number of other scientists, not mentioned by Bishop, have also made suggestions about the possible role of indeterminism in the brain in recent years, including, interestingly, its potential evolutionary significance. They include neuroscientist, Peter Ulric Tse, who has made detailed and highly original suggestions about these topics in a recent book (2013), as well as neuroscientists Paul Glimcher (2005) and Michael Shadlen (2014), biologists Bjorn Brembs (2011) and Martin Heisenberg (2013), astrophysicist Robert Doyle (2011), physicists G.F.R. Ellis (2009), John Polkinghorne (2009), David Layzer (2022), and psychologist Dean Simonton (2004), among many others.[4] It remains an open scientific question, of course, whether indeterminism does function in the neural processing of the brain. But rather than being dismissed

out of hand, as in the past, this possibility is now regarded as a serious one by these and other scientists.

Yet our question at present is a philosophical one that has boggled people's minds since the time of the ancient Stoics and Epicureans: What could one *do* with indeterminism, assuming it was there in nature in the right places, to make sense of free will as something other than mere chance or randomness and without appealing to mystery? If minute quantum indeterminacies in the firings of individual neurons were amplified so that they introduced some indeterminism into the larger-scale processing in deliberation and decision-making, how could this help to make sense of free choice as something other than mere chance? This is the Intelligibility Question, just defined, about an incompatibilist free will that we must now address.

## 17 Initial Pieces: Self-formation, Efforts, Willpower, Volitional Streams

Let us begin by recalling that indeterminism does not have to be involved in all actions done "of our own free wills." It need be involved only in those choices or acts by which we make ourselves into the kinds of persons we are, with the wills we have. These are the "will-setting" or "self-forming" actions (SFAs) of earlier sections.

I believe these SFAs would occur at those difficult times in life when we are torn between competing visions of what we should do or become; and they would be more common in everyday life than one may think. Perhaps we are torn between doing the moral thing or acting from ambition, or between present desires and long-term goals, or we are faced with difficult tasks for which we have aversions, or have to exert willpower to keep prior commitments and resolutions rather than break them. In all such cases and many others, we are faced with competing motivations and have to strive or make an effort or exert willpower to overcome the temptation to do something else we also strongly want.

At such times, the tension and uncertainty we feel about what to do, I suggest, would be reflected in some indeterminacy in our neural processes themselves – in the form of amplified background neural indeterminacy as described in Section 16 – neural indeterminacy that is "stirred up," one might say, by the conflicts in our wills. The uncertainty and inner tension we feel at such soul-searching moments of self-formation would thereby be reflected 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.

A further step would then involve noting that in such cases of self-formation, where we are faced with competing motivations, whichever choice is made will

require an effort of will or exercise of willpower to overcome the temptations to make the alternative choice. I thus postulate, in such cases, that different goal-directed cognitive processes ("volitional streams," we might call them) might be involved in the brain, corresponding to these exertions of effort or willpower. These cognitive processes or efforts would have different goals corresponding to the different choices that might be made (e.g. a moral choice or a self-interested choice). But, importantly, it is not being claimed that these efforts or exercises of willpower aimed at different choices would be occurring at the same time during deliberation. Nor would they be occurring throughout the entire deliberation. Rather, different efforts or exertions of willpower may be initiated at different times depending on the course of the agent's reasoning.

To illustrate, consider a familiar example of Peter van Inwagen's (1983) of a would-be thief, call him John, who is deliberating about whether or not to steal from a church poor box. Suppose John is deeply torn because, on the one hand, he is desperately in need of money and knows that no one is usually in the church on weekday afternoons, so he can likely steal the money without being caught. On the other hand, he has moral qualms about doing so because he knows that the money in the poor box is used to help other people who are in need, some of whom may need it as much as, or more than, he does.

We might then imagine that in the course of John's deliberation, various thoughts, experiences, and memories come to mind, various preferences, desires, and possibilities are assessed and weighed, so that his considered reasons incline him to choose to steal the money rather than not to steal it. Of crucial importance, however, if this is a self-forming choice situation in the sense described in prior sections, we must say that the reasons motivating the choice to steal the money merely *incline* John to make that choice at this time rather than the alternative choice. These reasons do not determine he will do so. To use a traditional expression of Leibniz, his reasons "incline without necessitating." If a choice is thus to be made in accord with these inclinations, effort would have to be made or willpower exerted to overcome the resistance in his will. This resistance would be coming from his motives to make the contrary choice, which motives also remain important to him.

This is where indeterminism would enter the picture as well. For, in the manner described earlier in this section, this conflict in John's will would "stir up" indeterminism in the effort to make the choice to which he is currently inclined (to steal from the poor box), making it uncertain the effort will succeed in attaining its goal. If the effort to choose to steal from the poor box in terms of his presently inclining reasons does succeed, despite this indeterminism, the choice to steal to which John is presently inclined would be made and the deliberation would terminate.

Note that if this should happen, the choice to steal, *despite being undetermined*, would have been made *by John* purposefully and in accordance with his

will. For it would have been the result of a goal-directed effort of will to make just this choice at this time rather than an alternative choice. Moreover, the choice would have been made for the reasons inclining him toward that choice rather than the alternative at the time. Thus, it wouldn't have been a mere accident that the choice occurred, *even though its occurrence was undetermined*. The choice would have been brought about voluntarily and on purpose, as a result of the goal-directed effort of the agent.

What would happen, however, if due to the indeterminism involved, the effort to choose to steal from the poor box did *not* succeed at that time and the choice had not been made? Many critics of a free will requiring indeterminism assume that if a choice is undetermined, the agent would be able to make a different choice (e.g. to steal or not to steal) given exactly the same deliberation leading up to moment of choice, including exactly the same desires, beliefs, thoughts, inclinations, and prior reasoning. And given this assumption, it would follow, these critics argue, that if John had failed to choose to steal from the poor box at the time he did choose, due to the indeterminism involved, he would instead have chosen *not* to steal from the poor box at that time instead. And this seems problematic, these critics argue, given that his deliberation would have been exactly the same leading up to the choice. What would explain the difference in choice?

But this commonly made assumption need not be made, nor is it made, in the account of self-forming choices being given here. It is not assumed, nor need it be assumed on this account, that if a choice is undetermined, the agent might make different choices (e.g. to steal or not to steal) given exactly the same deliberation, including exactly the same desires, beliefs, and reasoning, leading up to the choice. All that follows from the assumption that a self-forming choice or SFA is undetermined is that the effort to make it may succeed *or may fail* at a given time in overcoming the resistance in the will to making it. And from this, it does not follow that if the effort fails, an alternative choice would be made at that same time.

*Failure is rather a signal to the agent not to choose too quickly in terms of the presently inclining reasons. Failure says in effect: think more about this. The resistant motives for the alternative choice* (e.g. John's motives for *not* stealing from a poor box) *still matter to you and these resistant motives should not be dismissed too readily*. These resistant motives are the causal source of the indeterminism in the effort to choose to steal in the first place, making it uncertain that the effort will succeed here and now. The stronger these resistant motives are, the greater the probability the effort may fail, due to the indeterminism to which the resistant motives give rise.

In other words, a distinction needs to be made between John's *not choosing to steal* at a time and his choosing *not to steal* at that time. What is assumed, if John fails in his effort to choose to steal from the poor box at a time, due to the

indeterminism involved, is not that he would have made the contrary choice, not to steal, at that same time, but rather that no choice at all would have been made at that time. The deliberation might continue until a potential reassessment of the motivating reasons that inclined to one choice or the other led to another later effort to make the choice to steal or a potential reassessment led to a later effort to make the choice not to steal. Or, the deliberation might terminate without any decision being made.

John, we may imagine, if he fails to overcome the resistance in his will to making the choice to steal at a time, might reconsider his motivating reasons. Then, moved by his moral qualms about stealing money from a poor box used to help other people, he may be inclined seconds or minutes later to choose *not* to steal and make an effort to choose in accord with that inclination. The success of this further effort would also be undetermined, but if it succeeded nonetheless, the choice not to steal would be made. Or, John may find on reconsidering that he really needs the money and makes a further effort at a time later to overcome his moral qualms. This effort may in turn fail as well, but if it succeeds despite the indeterminism, he would make the choice to steal at this later time. Or, the deliberation may terminate without any decision being made. John may leave the church planning to think more about it, perhaps berating himself for his indecisiveness.

Note that in any of these possible scenarios, if John does succeed at a time in an effort to make one or another of the choices to which he is inclined at that time, he will have brought about the choice made and will have done so voluntarily and intentionally and for the motivating reasons that inclined him toward that choice at that time. For he would have succeeded in an effort whose goal was to make that very choice for those inclining (though not necessitating) reasons; and this would be the case even though the choice were undetermined.

> **Indeterminism** would have been involved in the effort, but it would not be the cause of the choice, if the effort succeeds. For the effort would have succeeded, *despite* the indeterminism and not *because* of it. The cause of the choice would have been the *agent*, whose effort or exercise of willpower brought about the choice.

Note also that the indeterminism that is ingredient in the agent's effort to make the choice to which the agent is then inclined *is not an accidental feature of the situation*. It does not just *happen* to be present. The presence of the indeterminism is rather a consequence of the conflict in the agent's will and of the resistant motives that are a feature of that conflict – resistant motives that have to be overcome by effort, whichever choice is made. The stronger these resistant motives are, the greater the degree of indeterminism stirred up, and the greater the probability of failure of the effort.

The idea is thus to *think of the indeterminism involved in self-forming choices*, not as a cause *acting on its own*, but as an *ingredient* in larger *goal-directed* activities of the agent, in which the indeterminism functions as a *hindrance* or *interfering* element in the attainment of their goals. The choices that result would then be *achievements* brought about by the goal-directed activities (the efforts of will or exercises of willpower) of the agent, which might have failed since they were undetermined, but one or the other of which might succeed in its goal.

Moreover, if such processes aimed at different goals may occur at different times in the course of deliberation (in the conflicted circumstances of a self-forming choice), *whichever choice may be successfully made will have been brought about by the agent's volitional striving* (the effort) to make that particular choice rather than the other at that time, despite the possibility of failure due to the indeterminism.

## 18 Indeterminism and Responsibility

Another significant consequence of thinking of indeterminism in this way is the following: when indeterminism functions in this manner as an obstacle to the success of goal-directed activities, the *indeterminism does not undermine responsibility*, *if the activities succeed in attaining their goals*, despite the indeterminism.

Consider the example introduced in Section 6 of an assassin trying to kill a prime minister from a distance with a high-powered rifle when, owing to a nervous twitch in his arm, he fails to hit his target. Or consider another example of a husband arguing with his wife, who in anger swings his arm down on her favorite glass tabletop intending to break it. In each of these cases we could imagine that an element of genuine chance or indeterminism is involved. We might imagine that the nervous twitches in the arms of the assassin that lead to missing his target – or the reduced momentum in the swing of the husband's arm that might lead to his failing to break the tabletop – are the result of undetermined quantum events in their brains or nerve pathways.

Due to this indeterminism in their nerve pathways, they might fail to do what they were intending and trying to do. But suppose that, despite the indeterminism, the assassin *succeeds* in his goal of killing the prime minister and the husband in breaking the tabletop. In such cases, both the assassin and the husband would be fully responsible for their actions, because both would have succeeded in doing what they were intending and trying, and making efforts to do, despite the indeterminism involved.

It would be a poor excuse for the assassin to plead in the courtroom that he was not guilty of killing the prime minister because due to the indeterminism

in his nerve pathways it was undetermined, and hence a matter of chance, that he succeeded in hitting his target. It would be equally absurd for the husband to offer the excuse to his wife that, since it was undetermined that his arm swing would break the tabletop, the breaking of the tabletop was a matter of chance *and so he was not responsible*. His wife would not be impressed – and for good reason.

For there was indeed a "chance" these agents would fail in doing what they were trying or making efforts to do. But if they succeeded, nonetheless, chance would not have been the *cause* of the prime minister's death or the table's breaking. *They*, the agents, would have been the causes, by virtue of the fact that they would have succeeded in doing what they were intending and trying, and making efforts, to do.

> When indeterminism thus functions as an obstacle to the success of goal-directed activities, the indeterminism does not preclude full responsibility, if the activities succeed in attaining their goals nonetheless.

This would be the case for self-forming choices or SFAs as well, but with an important difference. *Whichever choice the agents should make* in the course of a deliberation in a self-forming choice situation, the agents would have succeeded in doing what they were making an effort to do at that time, despite the indeterminism involved in their neural processing.

If John, for example, chooses to steal from the poor box at any time during his deliberation, it will be due to the success of his effort to make that choice at that time, thereby overcoming the resistance in his will to doing so. And if he chooses not to steal from the poor box at any other time in the deliberation, it will be due to the success of his effort to make that other choice not to steal at that time, thereby overcoming the resistance in his will to doing so. Whichever choice is made in such self-forming choice situations, the indeterminism involved would thus not be a cause acting on its own but an ingredient in a larger goal-directed cognitive activity of the agent that would have succeeded in attaining its goal, *despite* the indeterminism and not *because* of it. The agents would be responsible for the choice made since they would have succeeded in doing what they were intending and trying to do; and this would be the case whichever choice should be made in the course of the deliberation.

To sum up, I have been arguing that self-forming actions (SFAs) occur at those difficult times in life when we are torn between competing visions of what we should do or become. On such occasions of self-formation, agents are, as is often said, "of two minds." Yet they are not two separate persons. They are not dissociated from either of their conflicted states of mind.

Consider a young woman who is about to graduate with honors from a law school and who is deliberating about which of two attractive job offers to accept. One offer is with a large corporate law firm in a big city, the other a smaller,

up-and-coming, but less prestigious, firm in a smaller city near where she grew up. She is torn because each firm has features that are deeply attractive to her. On the one hand, people at the smaller firm are more friendly and there is more of a chance of attaining an eventual partnership there. The smaller firm is also near to where she grew up and hence to her family and many friends. On the other hand, she is extremely ambitious and the chance to be part of this very prestigious firm in a large city is very attractive to her, despite the difficulties it may involve. She is a small town person with big time ambitions.

The young woman of this example is a complex creature, torn inside by different visions of who she is and what she wants to be, as we all are from time to time. But **this is the kind of complexity needed for genuine *self-formation*** and hence for ***free will rather than merely freedom of action***. And when agents, like this young 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. They thereby **voluntarily and purposefully commit themselves to a particular pathway into the future; and this will be so, whichever choice they should succeed in making,** despite the indeterminism involved.

## 19 Initial Questions and Objections: Indeterminism and Chance

Many questions and potential objections naturally arise about this view as so far presented. Addressing them will allow us to bring out more features of the view, which is far from complete. Many of the most obvious objections people have to views of free will requiring indeterminism, including the view presented here, rest on intuitions they have that if choices are undetermined, they *must* happen merely by chance – and so must be "random," "capricious," "uncontrolled," "irrational," and all the other things usually charged. Such intuitions are deeply rooted. But if we are going to understand free will, I think we have to break habits of thought supporting such intuitions and learn to think in new ways.

The first step would be to question the intuitive connection in people's minds between "indeterminism's being involved in something" and "it's happening merely as a matter of chance or luck." "Chance" and "luck" are terms of ordinary language that imply "something's being out of one's control." So, using them already begs certain questions and may mislead us. Whereas:

> "Indeterminism" is a technical term that merely *rules out deterministic causation, not causation altogether. Indeterminism is consistent with probabilistic forms of causation*, where outcomes are caused, but not inevitably. It is thus a mistake – one of the most common mistakes in the long history of debates about free will – to assume that "undetermined" must mean or imply "uncaused."

Another common source of misunderstanding is the following: suppose our young law graduate does decide to join the larger firm in the big city. If her decision is undetermined up to the moment when it occurs, one may have the image of her first making the effort to overcome the still strong motives to do otherwise (to join the smaller firm) and then at the last instant "chance takes over" and decides the issue for her. But this image is misleading. On the view presented, one cannot separate the indeterminism and the effort, so that *first* the effort occurs *followed by* chance or luck. Rather, the efforts or exertions of willpower are temporally extended goal-directed processes of the agent and the indeterminism is an ingredient in these larger processes, not something separate that occurs *after* or *before* them. The neural networks that realize the efforts in the brain are circulating impulses and there is some indeterminacy assumed in the timings of firings of individual neurons in these circulating impulses. But these processes as a whole are her efforts and they persist right up to the moment when the decision is made.

*There is no point at which her effort stops and chance "takes over."* She decides as a causal result of her effort, even though she might have failed due to the indeterminism ingredient in the effort. Likewise, the husband breaks the table as a causal result of his effort, even though he might have failed because of the indeterminacy. This is why his excuse "chance broke the table, not *me*" is so lame when he succeeds.

## 20 Further Questions and Objections: Phenomenology and Rationality[5]

Yet another frequently made objection is that we are not introspectively or consciously aware of making efforts and performing multiple cognitive tasks in self-forming choice situations. But it is not being claimed that agents must always be introspectively aware of making such efforts or exertions of willpower (though sometimes they might be when they are very conflicted). And importantly, as emphasized in Section 5, it is not being claimed that these efforts or exercises of willpower 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 exercises of willpower may be initiated at different times depending on the course of the agent's reasoning.

What persons are introspectively aware of in self-forming choice situations is that they are trying to decide about which of two competing options to choose and either choice is difficult because there are resistant motives pulling them in different directions, some of which will have to be overcome, whichever choice is made. In such introspective conditions, I am theorizing that what

is going on underneath is complex processing in the brain that may involve, at various times in the course of deliberation, one or another goal-directed cognitive process whose goal is making a specific choice and thereby overcoming resistance in the will to making that choice.

There is a more general point here that I have often emphasized: *introspective evidence cannot give us the whole story about free will*. Stay on the conscious surface and libertarian free will is likely to appear obscure or mysterious, *as it so often has in history*. What is needed is a theory about what might be going on behind the scenes when we exercise such a free will, not merely a description of what we immediately experience. There is, in fact, a growing body of empirical evidence showing that in complex cognitive processes, such as practical reasoning, much of the processing that is going on occurs unconsciously.[6] Moreover, some of what occurs unconsciously may involve effort making or exercises of willpower (e.g. to access memories or associations or considerations that may have a bearing on a decision, to overcome temptations to suppress other information we may not want to think about, to resist strong inclinations, to avoid rationalizations or self-deception, and so on).

Another common objection is that it is irrational to make efforts to do incompatible things. I agree it would be irrational, if the efforts to make incompatible choices (say to steal or not to steal) were being made at the same time, given exactly the same reasoning up to that time. But this is not what is being assumed here in the case of self-forming choices. Rather one or another of these efforts or exercises of willpower may be initiated at different times, depending on the trajectory of the agent's reasoning up to that time. In particular, one such effort may be initiated when, in the course of deliberation, the agent's considered reasons at that point incline (without necessitating) the agent to make one of the choices rather than another. It is not irrational to make an effort to make a choice in such situations in terms of one's presently inclining reasons. Though it would be irrational to also make an effort to make an opposing choice at *this same* time, given these same inclining reasons.

It is important in this regard to recognize the uniqueness of such self-forming choice situations. For our normal intuitions about efforts are formed in everyday situations in which our wills are already "set one way" on doing something, where obstacles and resistance have to be overcome, if we are to succeed in doing it. We want to open a door which is jammed, so we have to make an effort to open it. Such ordinary situations are what were earlier called "will-settled situations," where our wills are already set or settled on doing what we are making efforts to do. I am making an effort to open a jammed closet door to get what is inside. There is no resistance in my will to doing so, no reasons to do otherwise: I need what is in the closet for my day's activities. The resistance that has to be overcome by effort thus has an external source, in the conditions of the door and the door frame. The resistance is not coming from my own will.

By contrast, self-forming choices or SFAs, as we have seen, are will-*setting*, not already will-*settled*. Our wills are *not* already settled on doing what we are making efforts to do. The resistance that has to be overcome by effort is thus not coming from an external source. It is coming from our own wills. We *set* our wills one way or the other only in the act of choosing itself, when an effort we are making succeeds in overcoming the resistance in our will to making the choice in question.

This feature of will-setting choices – that the resistance to making them is coming from our own will, not from an external source – is related in turn to the fact that the reasons motivating the efforts to make such will-setting or self-forming choices merely "incline" without necessitating. The reasons motivating an opposing choice, which still matter to the agent, must be overcome by effort, if the choice to which the agent is presently inclined is to be made. It is thus rational to make an effort in such circumstances in terms of one's presently inclining reasons, if the resistance in one's will is to be overcome. What would *not* be rational would be to make an effort to make a contrary choice at this same time, given these same inclining reasons.

It would also be irrational to make no effort at all to overcome the resistance in one's will to making the choice to which one is inclined, but rather to leave the outcome to "chance" and "hope" the choice to which one is inclined "wins" out.

> Self-forming choices are not a matter of certain motivations *winning out* over others *on their own*. Rather, self-forming choices involve the *agent bringing it about* that one set of motivations wins out over another, by making an effort to do so and succeeding in that effort.

Because most efforts in everyday life, such as the effort to open the jammed closet door, are made in will-settled situations where our wills are already set on doing what we are trying or making efforts to do, we tend to assimilate all effort-making to such situations. We thereby fail to consider the uniqueness of ***will-setting***, which is of a piece, in my view, with the uniqueness of **free will**.

## 21 Micro- vs. Macro-control

Another common line of reasoning lying behind many objections to an undetermined free will is the following: Is it not the case, one might ask, on the view proposed that whether agents succeed in making a choice A in the circumstances of a self-forming choice (i) depends on whether certain neurons involved in their cognitive processing fire or not (perhaps within a certain time frame), is it not the case that (ii) whether or not these neurons fire is undetermined and hence a matter of chance, and hence that (iii) the agent does not have control

over whether or not they fire? But if these claims are true, it seems to follow that the choice merely *happened* as a result of the chance firings of these neurons and so (iv) the agent would not have had control over whether the choice of A was made or failed to be made and (v) hence the agent would not be responsible for making the choice.

For many persons, this line of reasoning clinches the matter. It looks like the outcome *must* be merely a matter of chance or luck after all. Yet they reason too hastily. For the really astonishing thing is that even though agents do not have control over whether or not the undetermined neurons involved in their cognitive processing fire or not, it does not follow that the agents do not have sufficient control to be responsible for the choices ultimately made. This does not follow when the following three things are also true: (i) the choosing of A rather than B (or B rather than A, whichever occurs) was something the agent was striving or trying to bring about at the time, (ii) the indeterminism in the neuron firings involved in this striving or trying was a hindrance or obstacle to the achievement of that goal, and (iii) the agent's striving or trying nonetheless succeeded in achieving the goal despite the hindering effects of the indeterminism.

For, consider the husband swinging his arm down on the table. It is also true in his case that (i) whether or not his endeavoring or trying to break the tabletop succeeds *depends* on whether certain neurons in his nervous system fire or do not fire. It is also true in his case that (ii) whether these neurons fire or not is undetermined and hence a matter of chance and is (iii) not under his control. Yet, even though we can say all this, it does not follow that (iv) the husband did not break the tabletop, and that (v) he is not responsible for breaking the tabletop, if his endeavoring or trying to do so succeeds. And, importantly, each of these things would be true in the case of a self-forming action, whichever choice should be made. Astonishing indeed! Yet this is the kind of surprising result one gets when indeterminism or chance plays an interfering or hindering role in larger goal-directed activities of agents that may succeed or fail.

It is well to meditate on this: we tend to reason that if an action (whether an overt action of breaking a table or a mental act of making a choice) depends on whether certain neurons fire or not (in the arm or in the brain), then the agent must be able to make those neurons fire or not, if the agent is to be responsible for the action. In other words, we think we have to crawl down to the place where the indeterminism originates in the individual neurons and makes them go one way or the other. We think we have to become originators at the micro-level and "tip the balance" that chance leaves untipped, if we (and not chance) are to be responsible for the outcome. And we realize we cannot do that. But we do not have to. It is the wrong place to look. We do not have to micromanage our individual neurons one by one to perform purposive actions and we do not have such *micro-control* over our neurons even when we perform ordinary free actions such as swinging an arm down on a table.

What is needed when we perform purposive activities, mental or physical, is *macro-control* of processes involving many neurons – processes that may succeed in achieving their goals despite the interfering or hindering effects of some recalcitrant neurons.

We do not have **micro-control** our each individual neuron or muscle that might be involved in our purposive activities. But that does not prevent us from having **macro-control** over these purposive activities themselves (whether they be mental activities such as practical reasoning, or physical activities, such as arm-swingings) and being responsible when those purposive activities attain their goals.

## 22 Control and Responsibility

But if indeterminism does not take away control altogether, does it not at least *diminish* the control agents have over their actions? Is it not the case that the assassin's control over whether the official is killed (his ability to realize his purposes or what he is trying to do) is lessened by the undetermined impulses in his arm – and so also for the husband and his breaking the table? Their control is indeed lessened. But a further surprising thing worth noting is that *diminished control in such circumstances does not entail diminished responsibility*, when agents *succeed* in doing what they are trying or making efforts to do. The assassin is not less guilty of killing the official if he did not have complete control over whether he would succeed because of the indeterminism; nor is the husband less guilty of breaking the table if he succeeds, despite the indeterminism involved.

Suppose there were three assassins, each of whom killed an official. Suppose one of them (an older assassin contemplating retirement) had a 50% chance of succeeding because of the indeterministic wavering of his arm, another had an 80% chance, and the third (a young stud) nearly a 100% chance. Is one of these assassins less guilty than the others, *if they all succeed*? It would be absurd to say that one assassin deserves a hundred years in jail, the other eighty years, and the third fifty years. The diminished control in the assassins who had an 80% or a 50% chance does not translate into diminished responsibility when they succeed.

There is an important further lesson here about free will in general. We should concede that indeterminism, wherever it occurs, does diminish control over what we are trying to do and is a hindrance or obstacle to the realization of our purposes. But recall the case of the young law graduate mentioned in Sections 18 and 19. The indeterminism that is admittedly diminishing her control over the choice she may be trying to make (to join one law firm or another) is coming from her own will. It is coming from the motives she has for making the opposing choice (to join the competing firm).

In each case, the indeterminism is functioning as a hindrance or obstacle to her realizing one of her purposes – a hindrance or obstacle in the form of resistance within her will which has to be overcome by effort. If there were no such hindrance – if there were no resistance in her will – she might indeed in a sense have "complete control" over one of her options. There would be no competing motives standing in the way of her choosing it and therefore no interfering indeterminism. But then also, she would not be free to *rationally* and *voluntarily* choose the other purpose (choose otherwise), because she would have no good competing reasons to do so. Thus:

> By being a hindrance to the realization of some of our purposes, indeterminism opens up the genuine possibility of pursuing other purposes – of choosing or doing otherwise in accordance with, rather than against, our wills. *To be genuinely self-forming agents (creators of ourselves) – to have free will – there must at times in life be obstacles and hindrances in our wills of this sort that we must overcome*. Free will is a gift, but it also involves struggle – and achievement.

## 23 Agency, Complexity, Disappearing Agents

Another question that has had a hypnotic effect on modern free will debates, reflecting deeply rooted intuitions, is the following: Do we not have to postulate an additional kind of "agent-causation" over and above causation by states and events to fully capture libertarian free choices, given that such choices must be undetermined by prior states and events? There is a residual fear underlying questions of this kind that the "agent" will somehow "disappear" from the scene if we describe its capacities and their exercise, including free will, in terms of causation by states and events alone involving the agent. I believe this fear to be ultimately misguided.

A continuing substance (e.g. an agent) does not absent the ontological stage because we describe its continuing existence – its life, if it is a living thing – including its capacities and their exercise, in terms of states, events, and processes involving it. One needs more reason than this *to think that agents do not cause things, only events cause things*. Human agents are continuing substances with both mental and physical properties. But it is not inconsistent to say this and to say that the *lives* of agents, their capacities, and the *exercise* of those capacities, including free will, must be spelled out in terms of states, processes, and events involving them. In short:

> One does not have to choose between **agent (or substance) causation** and **event causation** in describing freedom of choice and action. One can affirm both.

In the case of self-forming choices or SFAs, for example, it is true to say both that "the agent's deliberative activity, including their effort, caused or brought about the choice" and that "the agent caused or brought about the choice." Indeed the first claim *entails* the second. Such event descriptions are not meant to deny that agents, qua substances, cause their free choices and free actions. Rather, the event descriptions spell out in more detail *how* and *why* the agents did so. There is thus no reason to worry that the "agent" will somehow "disappear" from the scene if we describe its capacities and their exercise, including free will, in terms of causation by states and events involving the agent.

Relevant here to explaining the role of agents in the causation of action is a peculiarly modern scientific way of understanding this role that has roots in ancient views, such as that of Aristotle.

> **Agents**, according to this modern conception with ancient roots, are to be conceived as ***information-responsive complex dynamical systems***. "An agent's causing an action" is to be understood as "an agent, conceived as such an information-responsive complex dynamical system, exercising ***teleological guidance control*,** over some of its own processes."

*Complex dynamical systems* are understood in this context in the manner of "dynamical systems theory." Such systems (now known to be ubiquitous in nature and which include living things) are systems in which emergent capacities arise as a result of greater complexity. When the emergent capacities arise, the systems as a whole impose novel constraints on the behavior of their parts.

Such complex systems exhibit **teleological guidance control (TGC)** when they tend through feedback loops and error correction mechanisms to converge on a goal (called an attractor) in the face of perturbations.

Such control, as neuroscientist Marius Usher argues (2006), *is necessary for any voluntary activity* and he interprets it in terms of dynamic systems theory, as I would as well. Neuroscientists E. Miller and J. Cohen (2001) argue that such cognitive (guidance) control in human agents stems from the active maintenance of patterns of activity in the prefrontal cortex that represent goals and the means to achieve them. These patterns provide signals to other brain structures, they argue, whose net effect is to guide the flow of activity along neural pathways that establish the proper mappings between inputs, internal states, and outputs (Miller and Cohen 2001: 403).

An important consequence of understanding the agent causation involved in free agency and free will in this way is that the causal role of the agent in intentional actions of the kind needed for free agency and free will is not *reducible* to causation by mental states of the agent alone, such as beliefs, desires, and intentions. That would leave out the added role of the agent, qua complex dynamical *system*, exercising teleological guidance control (TGC)

over the processes *linking* mental states to actions. In the absence of this *systemic control* by the agent over the *manner* in which the mental states cause the resulting events, the causation by mental states might be "deviant" and the outcomes would not be intentional actions of the agent.

A further significant consequence of understanding causation of free actions in this way, as neuroscientist Usher (2006) points out, is that while the TGC of the kind required is compatible with determinism, it is also compatible with indeterminism. A complex dynamical system can exhibit TGC, tending through feedback loops and error correction to converge on a goal, even when, due to the presence of indeterminism, it is uncertain whether the goal will be attained. Such control is necessary for any voluntary activity, as noted, whether the voluntary activity is will-settled and determined or will-setting and not determined.

To sum up, one does not have to choose between agent (or substance) causation and event causation in accounting for free agency, libertarian or otherwise. You can, indeed you must, affirm both. And the agent or substance causation involved is not reducible to event causation by mental states alone for the reasons explained. There is thus no "disappearing agent problem" as well.

My view has often been called an event-causal (EC) libertarian view of free will to distinguish it from the two other familiar kinds of libertarian views, agent-causal (AC) and noncausal (NC). But I was never happy with this EC designation because of what has been said in this section, namely that an adequate libertarian theory of free will must appeal to both agent-causation and event-causation, and neither is reducible to the other. Thus, I now argue that we need to add to the three standard kinds of libertarian views (AC, EC, and NC) a fourth kind of view which I call AC/EC (if that doesn't sound too much like a rock group, as I noted when this title was first introduced), a view that requires both agent and event causation. My view, I argue, has always been of this AC/EC kind, which I now refer to as "a fourth way forward" in understanding libertarian free will.

## 24 Regress Objections: Responsibility and Character Development

Another common worry is that views of free will requiring UR lead to a vicious regress. To be ultimately responsible for a choice that issues from an agent's present will (character, motives, and purposes), the agent must be at least in part responsible by virtue of choices or actions voluntarily performed in the past (SFAs) for having the will he or she now has. But to be ultimately responsible for these earlier SFAs by which we formed our present wills, would we not have to be responsible in turn for the characters, motives, and purposes from

which these earlier SFAs issued? And would this not require still earlier SFAs by which we formed these characters, motives, or purposes?

We would thus be led backwards to the earliest choices of childhood when the wills from which we chose were not formed by us at all but were entirely the product of influences outside ourselves: parents, social conditioning, heredity, genetic dispositions, and so on. It may thus appear that all responsibility for later choices in life would go back to the earliest choices of childhood when we seem to have far less freedom and responsibility than we have later in life, which is absurd.

The first response to make to this familiar worry is to note that UR for choices in later life need not have its source entirely in choices of childhood. This would be true only if we made no subsequent self-forming choices in later life. Whereas, by contrast, the account of self-forming choices given in Sections 4–7 implies that if self-forming choices are possible for agents at all, they would normally occur throughout our lives and more so as we mature and life becomes more complex. This is so because:

> It is the complexity of our lives, and of our wills and motivations, that gives rise to conflicts in our wills and to **self-forming actions (SFAs)** in the first place; *and this complexity does not abate, but normally grows, as we develop beyond childhood*. In making self-forming choices as we mature, we would be constantly forming and *reforming* our existing characters, motives, and purposes as we go along in ways that, while influenced by our prior characters, motives, and purposes, are not determined by our prior characters, motives, and purposes.

I argue therefore, in partial agreement with philosophers, such as Aristotle, who talk about the development of "character," that responsibility for our *wills* (characters, motives, and purposes) accumulates over time (Kane 1996). Putting the matter in terms of the present theory: by making many self-forming choices through a lifetime, we gradually form and reform our characters, motives, and purposes in ways not determined by our past. It would follow that with regard to most of the self-forming choices we make, our responsibility has a twofold source: first, in the self-forming choices themselves we make in the present between our conflicting motives and purposes and, second, in the conflicting motives and purposes themselves from which the choices are made, many of which had *their* source in earlier SFAs by which we gradually formed our present wills over time.

The only exceptions, of course, would be the very earliest SFAs of childhood when it *is* normally true, if we go back far enough, that the motivations among which we choose all come from sources outside ourselves, parents, society, upbringing, etc. I have discussed these first SFAs of childhood in a number of writings (e.g. Kane 2011) and have a distinctive view about them, which may be spelled out as follows.

In the earliest SFAs of childhood, our responsibility, so far from being the source of all later responsibility, *is* very limited, precisely because there is as yet no backlog of self-formed character. That is why we hold children less responsible the younger they are. I further argue that:

> **The earliest self-forming actions (SFAs) of childhood** have a probative (or probing or learning) character to them. Young children are often testing what they can get away with and what consequences their behavior will have on them and others (among the many reasons why child-rearing is so exhausting). Their character is thus slowly built up by how they respond to the responses to these earliest probes. *Character and purposes to which they commit themselves accumulate and they become more responsible for subsequent acts that flow not just from present efforts but from past formed character and purposes as well.*

If a three-year-old is told not to take more than his share of cookies, but tries to do so anyway the next time, resisting his conflicting motives not to disobey his parents, then the child is responsible. But he is not as responsible as when he does it a second, third or fourth time and it becomes a pattern of behavior. The wise parent will not punish him severely the first time, but may do so mildly, by withholding something he wants. *But the wise parent will also know that it is a mistake never to hold the child responsible at all for these earliest probes. For it is only by being so held responsible in however limited ways in our earliest years that we gradually become self-forming beings with wills of our own making.*

## 25 The Explanatory Luck Objection: Authors, Stories, Value Experiments, and *Liberum Arbitrium*

These reflections lead to one of the most common and powerful variants of the luck objection made against this view and many other libertarian theories during the past three decades. This objection, which has been called the "Explanatory Luck Objection," is stated in the following way by Alfred Mele (1998), one of its most astute and persistent defenders:

> **The Explanatory Luck Objection**: If different free choices could emerge from the same past of an agent, 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. the agent's choosing one thing rather than another) would therefore be just a matter of luck.

This objection in various formulations is now so widely cited and affirmed by critics of libertarian views of free will that it is often referred to as *the* luck objection in the literature. And many philosophers assume it is decisive. I think

they are mistaken. But I also think the objection has the power it has because it teaches us something important about free will.

The first obvious thing to be said in response to this luck objection is the following: in the case of self-forming choices as described here, it is not true to say, as the objection does, that "different free choices could emerge from the same past of an agent." This is not true, if it means the agent could make opposing choices (e.g. to steal or not to steal) given exactly the same prior reasoning leading up to the moment of choice. All that follows, as argued in Sections 5 and 6, from the fact that a self-forming choice is undetermined is that it might be made at a given time or might fail to be made at that time. It does *not* follow that the opposing choice (not to steal) would be made at that same time, given exactly the same reasoning leading up to the choice to steal. And this would be true whichever choice is made in a self-forming choice situation.

Moreover, for whichever choice should be made in a self-forming choice situation, (i) the agent would have caused or brought about that choice by succeeding in an effort to bring it about, thereby overcoming resistance in the will to doing so, (ii) the agent would have knowingly made that choice *rather than* the alternative, and (iii) the agent would have had the power to bring about the choice made and would have successfully exercised that power when it was made. This power was not unlimited since the effort through which it was exercised might have failed due to the indeterminism involved. But if the effort succeeded, the agent's power to make the choice would have been successfully exercised. (iv) The choice would have been made for reasons that inclined (without necessitating) the agent to make that choice at that time rather than the alternative – reasons that the agent then and there chose to act upon. (v) The agent would have made the choice rather than the alternative *voluntarily* (without being coerced against his or her will) and (vi) would have done so *intentionally* or on purpose, not merely by mistake, by succeeding in an effort aimed at making that very choice rather than the alternative.

If saying "the agent's choosing one thing rather than the another is just a matter of luck," as this explanatory luck objection also does, is meant to deny any of these things (i–vi) about such self-forming choices, then saying that the outcome was just a matter of luck seems to be the wrong conclusion to draw. And if one were to say that "just a matter of luck" is meant to be consistent with all of these things, the argument from luck would seem to lose much of its traction.

Ah, but not quite all traction; and this is where things get interesting. With powerful arguments in philosophy, it is not enough to show their conclusions do not necessarily follow from their premises. One needs also to show why they seem to have such power and seem irrefutable. The luck objection in this popular form does not show that libertarian free choices must be "just a matter of luck," if that entails denying any of the claims (i–vi). But it does show that there

is something to the oft-repeated charge that such self-forming choices must be *arbitrary* in a certain sense.

A residual arbitrariness seems to remain in all self-forming choices (SFAs) because the agents cannot in principle have *sufficient* or *overriding* (*conclusive* or *decisive*) prior reasons for making one option and one set of reasons prevail over the other. Therein lies the truth in this explanatory luck objection: *an undetermined free choice cannot be completely explained by the entire past, including past causes or reasons*; and I think it is a truth that reveals something important about free will. I have argued elsewhere that such arbitrariness relative to prior reasons tells us 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 (Kane 1996: 145–146). 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. Whether it is the right choice, time will tell. Meanwhile, I am willing to take responsibility for it one way or the other."

Of special interest here, as I have often noted, is that the term "arbitrary" comes from the Latin *arbitrium*, which means "judgment" – as in *liberum arbitrium voluntatis* ("free judgment of the will" – the medieval designation for free will). Imagine a writer in the middle of a novel. The novel's heroine faces a crisis and the writer has not yet developed her character in sufficient detail to say exactly how she will act. The author makes a "judgment" about this that is not determined by the heroine's already formed past, which does not give unique direction. In this sense, the judgment (*arbitrium*) of how she will react is "arbitrary," but not entirely so. It had input from the heroine's fictional past and in turn gave input to her projected future. In a similar way:

> Agents who exercise free will are both authors of and characters in their own stories at once. By virtue of "self-forming" judgments of the will (*arbitria voluntatis*) (SFAs), they are *"arbiters" of their own lives, "making themselves" out of a past that, if they are truly free, does not limit their future pathways to one*.

If we should charge them with not having sufficient or conclusive prior reasons for choosing as they did, they might reply:

> True enough. But I did have "good" 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 I still am not, for that matter). *Like the author of the novel, I am in the process of writing an unfinished story and forming an unfinished character who, in my case, is myself.*

## 26 Contrastive Explanations

Closely related to this explanatory luck objection is another objection concerning "contrastive explanation" that is frequently made against theories of free will requiring indeterminism. A contrastive explanation is an explanation for why one thing occurred *rather than* another. In the case of free choices, it would be an explanation in terms of an agent's prior character, reasons, or motives for why the agent made one choice rather than another.

The objection in this case is that if a self-forming choice (e.g. between A and B) is undetermined up to the moment when it is made, there could be no adequate contrastive explanation for why it was made rather than the alternative choice. For the fact that the choice was undetermined would mean that either choice (of A or of B) might have occurred, given the totality of the agent's traits of character, motives, and reasoning preceding the moment of choice. And there thus could not be an explanation for why one choice was made rather than the other at that moment in terms of the totality of the agent's character, motives, and reasoning prior to choice.

The first thing to be said in response to this familiar argument is similar to the first thing said in response to the explanatory luck objection: in the case of self-forming choices as understood here, it is not true to say, as this objection does, that either choice (of A or of B) might have occurred, "given the totality of the agent's traits of character, motives, and reasoning preceding the moment of choice." All that follows from the fact that a self-forming choice (e.g. the choice of A) is *undetermined* at a given time is that it might be made at that time or might *fail* to be made at that time. It does *not* follow that if the choice (of A) fails to be made at that time, the opposing choice – (of B) – would be made *at that same time*, given exactly the same reasoning that led to the choice of A.

But those who make this objection concerning contrastive explanation to views of free will requiring indeterminism usually have another assumption in mind that also needs to be addressed. They often assume that for an explanation of a free choice to be adequately contrastive in the sense they require, the following would have to be the case: if making the choice that was made during a deliberation rather than any alternative was the rational or reasonable thing to do, given the totality of the agent's reasons or motives, then making an alternative choice during that same deliberation, given the totality of the agent's reasons or motives, would *not* have been rational or reasonable. If making an alternative choice in the circumstances might also have been a rational or reasonable thing to do, we would not have an adequate contrastive explanation, in the sense these critics would require, for why one choice was made *rather than* another in terms of the agent's reasons and motives.

But if this is what contrastive explanations would require, there clearly could not be contrastive explanations in the sense these critics require of self-forming

choices or SFAs. For it is an essential feature of self-forming choices that no *such* strong contrastive explanations could be given for them. The reason is that, in addition to being undetermined, self-forming choices must satisfy *plurality conditions* for free choice: the power to make them and the power to do otherwise, *either way*, voluntarily, intentionally, *and rationally*. And this rules out the requirement that any other choice that might have been made in the course of a deliberation, other than the choice actually made, would have been unreasonable or irrational.

Moreover, this feature is not a defect of self-forming choices, according to the account given of them, but it is a consequence of their power. For it is precisely because agents have the power to make such choices and the power to do otherwise, voluntarily, intentionally, and rationally either way that makes it possible for such choices to be *will-setting* rather than *will-settled*. And the power to make will-setting choices at some points in our lives is what makes it possible for us to be makers or creators to some degree of our own wills rather than to be always acting from wills already formed.

It is also important to emphasize, however, that while agents who make such will-setting or self-forming choices may not have conclusive or decisive reasons for making the choice that is made rather than any other, such agents do nonetheless have reasons for choosing as they do that are "good enough" to render the choices they do make reasonable and rational ones, given their total reason sets when they choose. Some mathematical decision theorists speak in this connection of:

> **Satisficing reasons**: reasons that are good enough to justify a choice or action even though they are not sufficient to render any possible alternative choice or action that might have been made in the circumstances unreasonable or irrational.

Reasons for will-setting or self-forming choices are satisficing reasons in this decision-theoretic sense.

Moreover, the fact that the reasons for self-forming choices (SFAs) are satisficing in this sense is related to something important about free will that was spelled out in Section 25. It is related to the fact 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."

This "narrative" conception of self-formation, as we might call it, is nicely captured in an important recent book by John Doris *Talking to Ourselves: Reflections, Ignorance, and Agency* (2015a). In a section of this book in which

Doris talks about my views of agency and responsibility, he notes that in my defense of libertarian free will, I "develop the intriguing suggestion that ambivalence" about what one's true values are "and its resolution in action" is not contrary to responsible agency, but is essential to responsible agency (Doris 2015a: 162). It is so at least at some points in our lives when we are torn between conflicting values.

At such times, Doris says, when on my view we are engaged in self-formation, it is possible that more than one path into the future could represent our "true values," and it would be "up to us" which path we will take. We decide then and there which of our *possible* true values our actions will express. If we were never ambivalent – ambi-valent, I would say – in this way, we could not be self-creating beings, since our choices and actions would always be expressing what we *already* were, the formed will we already had.

At the point in his book, where Doris references these views of mine on conflicting values and ambivalence, he also discusses the example of Huckleberry Finn – an example that has played such a prominent role in contemporary philosophical writings on agency and ethics. On one telling of the Huck story, Doris says, "Huck held values favoring *both* the conventional course of action," that he should turn his friend and companion Jim, a black man who had escaped from slave owners, over to the authorities and, on the other hand, "the course [Huck] actually followed," of not turning Jim over. In sum, Doris says, "Huck's values *conflicted* . . . he suffered a kind of *ambivalence*" (Doris 2015a: 161).

That, I believe, is how Mark Twain himself tells the story. As I would put it:

> Huck is growing and developing as a self or agent. In deciding not to turn Jim in, Huck is not merely *expressing* what sort of a self he already is; he is also *deciding* what sort of a self he is going to be by deciding from among the conflicting values he has, which ones he will follow. He is thereby not merely engaged in self-*expression*, but in self-*making*, of the kind I believe *freedom of will* and not mere *freedom of action* sometimes requires. Such conflicts in the will and their resolution or lack thereof (as my wife, a writer, continually reminded me) are the stuff of most great literature and drama, Huck Finn, Madame Bovary, Hamlet, Anna Karenina, you name it.

## Further Reading

For a more advanced discussion of the issues considered in this chapter by various authors, see the collection of readings in two editions of Robert Kane (ed.), *The Oxford Handbook of Free Will* (Oxford, 2002, 2011). The libertarian view of free will presented in this chapter is further developed in the following books of mine: *Free Will and Values* (Paragon, 1985), *The Significance of Free Will* (Oxford, 1996), *A Contemporary Introduction to Free Will* (Oxford, 2005), and a

new, as yet unpublished work, *The Complex Tapestry of Free Will: A Philosophical Odyssey* (Oxford, accepted for publication) as well as numerous articles. A festschrift involving ten authors discussing my views of free will and my responses is edited by David Palmer, *Libertarian Free Will: Contemporary Debates* (Oxford, 2014).

Other libertarian views of free will, AC, EC, or (NC), different in various ways from the AC/EC view defended here, include Peter van Inwagen, *An Essay on Free Will* (Oxford, 1983), Timothy O'Connor, *Persons and Causes* (Oxford, 2000), Randolph Clarke, *Libertarian Accounts of Free Will* (Oxford, 2003), Carl Ginet, *On Action* (Cambridge, 1990), Hugh McCann, *The Works of Agency* (Cornell, 1998), Mark Balaguer, *Free Will as an Open Scientific Problem* (MIT, 2010), Laura Ekstrom, *Free Will: A Philosophical Study* (Westview, 2000), Stewart Goetz, *Freedom, Teleology and Evil* (Continuum, 2008: 303–316), David Hodgson, *Rationality+Consciousness = Free Will* (Oxford, 2012), E.J. Lowe, *Personal Agency: The Metaphysics of Mind and Action* (Oxford, 2008), Helen Steward, *A Metaphysics of Agency* (Oxford, 2012), Robert Doyle, *Free Will: The Scandal of Philosophy* (I-Phi Press, 2011), Kevin Timpe, *Sourcehood and its Alternatives* (Continuum, 2008), James Felt, *Making Sense of Our Freedom* (Cornell, 1994), Thomas Pink, *Free Will: A Short Introduction* (Oxford, 2004), Tim Mawson, *Free Will: A Guide for the Perplexed* (Continuum, 2011), and Meghan Griffith, *Free Will: The Basics* (Routledge, 2013). Three authors developing unique libertarian views are Christopher Evan Franklin, *A Minimal Libertarianism: The Promise of Reduction* (Oxford, 2018), John Lemos, *A Pragmatic Approach to Libertarian Free Will* (Routledge, 2018) and *Free Will and Values: Criminal Justice, Pride and Love* (Routledge, 2023), and Ken Levy, *Free Will, Responsibility and Crime* (Routledge, 2020) and "On Three Arguments Against Metaphysical Libertarianism" (*Review of Metaphysics* 76(4), 2023: 725–748). Alfred Mele, in his book *Free Will and Luck* (Oxford, 2006) and other works, develops some original possible libertarian views, without endorsing any particular view.

Useful introductions to scientific views of complexity and complex systems include R. Lewin, *Complexity: Life at the Edge of Chaos* (MacMillan, 1992) and M. Mitchell Waldrop, *Complexity: The Emerging Science at the Edge of Chaos* (Simon and Schuster, 1992). Works that attempt to apply theories about complex systems to issues of action and agency include E. Thelen and R. B. Smith, *A Dynamic Systems Approach to the Development of Cognition and Action* (MIT, 1994) and Alicia Juarrero, *Dynamics in Action: Intentional Behavior as a Complex System* (MIT, 1999). Writings by scientists suggesting possible roles for indeterminism in the brain and behavior include, among others, neuroscientist Peter Ulric Tse, *The Neural Basis of Free Will* (MIT, 2013), neuroscientist Paul Glimcher, "Indeterminism in Brain and Behavior," *Annual Review of Psychology* 56: 25–56), neuroscientist Michael Schadlin, "Comments on Adina Roskies: Can Neuroscience Resolve

Issues about Free Will?" in *Moral Psychology: Volume 4*, ed. by Walter Sinnott Armstrong (MIT, 2014: 139–150), biologist Martin Heisenberg "The Origin of Freedom in Animal Behavior," in A. Suarez and P. Adams (eds.), *Is Science Compatible with Free Will?* (Springer, 2013), physicist G.F.R. Ellis, "Top-down Causation and the Human Brain," in N. Murphy, G.F.R. Ellis, and T. O'Connor (eds.), *Downwards Causation and the Neurobiology of Free Will* (Springer, 2009), astrophysicist David Layzer, *Why We are Free: Consciousness, Free Will and Creativity in a Unified Scientific Worldview* (I-Phi Press, 2021), John Park, "Decision-making and Quantum Mechanical Processes of Cognitive Processing," *Journal of Cognitive Science* (accepted for publication), J.R. Busemeyer and P.D. Bruza, *Quantum Models of Cognition and Decision* (Cambridge University Press, 2012), Antonella Corradini and Uwe Meixner (eds.), *Quantum Physics Meets Philosophy of Mind: New Essays on the Mind-Body Relation in Quantum Theoretical Perspective* (De Gruyter, 2014), and Antoine Suarez and Peter Adams (eds.), *Is Science Compatible with Free?: Exploring Free Will and Consciousness in the Light of Quantum Physics and Neuroscience* (Springer, 2013).

## Notes

1. Libertarianism *about free will* should not be confused with political and economic doctrines of libertarianism. Libertarians about free will can, and do, hold differing views on political and economic matters.
2. I have addressed these issues in Kane (1985, 1996, 2005, 2007, 2011, 2014), among other writings.
3. Galen Strawson has made this case most forcefully with his "Basic Argument" (1986).
4. For example, Satinover (2001), Vasiri and Plenio (2010), Rolls (2012), Stapp (2007), Hameroff and Penrose (1996).
5. Objections addressed in this section and subsequent ones have been made in various forms by many critics of these features of libertarian views of free will, including Pereboom (2001), Clarke (2003), Mele (2006), Haji (2009), Levy (2011), Caruso (2012), among others. Other attempts to answer such objections have been made by defenders of libertarian views, including Balaguer (2010), Franklin (2018), Lemos (2018).
6. See, for example, Nichols (2015), Balaguer (2010), Jedlicka (2014), Usher (2006), Miller and Cohen (2001), Glimcher (2005), Shadlen (2014), and Brembs (2011).