

FREE WILL AND THE DIALECTIC OF  
SELFHOOD: CAN ONE MAKE SENSE OF A  
TRADITIONAL FREE WILL REQUIRING  
ULTIMATE RESPONSIBILITY?\*

*El libre albedrío y la dialéctica de la individualidad:  
¿se le puede dar sentido al libre albedrío tradicional  
que requiere de la responsabilidad última?*

ROBERT KANE\*\*

Department of Philosophy  
College of Liberal Arts

The University of Texas at Austin - U. S. A.

**ABSTRACT**

For four decades, I have been developing a distinctive view of free will according to which agents are required to be ultimately responsible for the creation or formation of their own wills (characters and purposes). The aim of this paper is to explain how a free will of this traditional kind—which I argue is incompatible with determinism—can be reconciled with modern developments in the sciences and philosophy. I address criticisms that a nondeterminist free will of this kind does not allow sufficient agent control, reduces to mere chance or randomness, fails to account of moral responsibility, and cannot be reconciled with modern science; and I relate such a free will to the nature of the self or person by developing what I call a “dialectic of selfhood.”

*Keywords:* Free will, dialectic of selfhood, ultimate responsibility, determinism, luck.

**RESUMEN**

Por cuatro décadas he venido desarrollando un enfoque particular del libre albedrío, de acuerdo con el cual éste requiere que los agentes sean últimamente responsables de la creación o formación de su propia voluntad (su carácter y sus propósitos). El objetivo de este artículo es explicar cómo una noción del libre albedrío de este tipo particular—de la cual sostengo que es incompatible con el determinismo—puede ser reconciliada con los desarrollos modernos en las ciencias y la filosofía. Enfrento las críticas según las cuales el libre albedrío no determinista no permite que haya suficiente control por parte del agente, se

---

\* The paper is a revised and expanded version of a previously unpublished lecture delivered at the Metanexus Institute Conference on Science, Philosophy and Religion in Madrid, Spain on July 12-17, 2008.

\*\* [rkane@uts.cc.utexas.edu](mailto:rkane@uts.cc.utexas.edu)

reduce a mera suerte o azar, falla en dar cuenta de la responsabilidad moral, o no puede ser reconciliado con la ciencia moderna; y relaciono este libre albedrío con la naturaleza del yo o la persona, al desarrollar lo que he llamado una “dialéctica de la individualidad”.

*Palabras clave:* libre albedrío, dialéctica de la individualidad, responsabilidad última, determinismo, suerte.

### 1. Modernity and Free Will

“There is a disputation that will continue till mankind is raised from the dead, between the necessitarians and the partisans of free will.” These are the words of the 12th century Persian poet and Sufi thinker, Jalalu’ddin Rumi. The free will of which Rumi speaks is the traditional notion of freedom that many thinkers have believed was in conflict with necessitarian or deterministic doctrines of all kinds —fatalistic, theological, physical, biological, psychological or social. Many centuries after Rumi, we are still debating about this notion of free will, whether we have it, whether it is or is not compatible with determinism, why it is thought by so many to be crucial to our sense of selfhood or personhood, how it is related to notions such as autonomy, rationality, responsibility, desert, dignity, morality, creativity, and others, that are thought to be crucial to our self-image as humans.

But while the debate about free will goes on in modern times, there are important new changes in recent debates on the subject, new directions taken and worth exploring. I want to discuss some of these new directions in this paper that are particularly related to my own work on free will over the past forty years.<sup>1</sup> The traditional idea of free will of which Rumi speaks —and which I believe to be incompatible with determinism— has been under sustained attack in modernity as outdated, obscure and unintelligible and has been dismissed by many modern philosophers and scientists since the 17<sup>th</sup> century for its supposed lack of fit with the modern images of the human beings and the cosmos in the natural and human sciences. Nietzsche summed up a prevailing view in his inimitable prose when he said:

The desire for “freedom of the will” in the superlative metaphysical sense [...] the desire to bear the ultimate responsibility for one’s actions oneself [...] to be nothing less than a *causa sui* [...] is the best self-contradiction that has been conceived so far [by the mind of man]. (2002 21)

---

<sup>1</sup> This work includes several books, notably Kane (1985, 1996, 2005), several edited volumes with included essays, notably Kane (2002a, 2002b), and many articles, including Kane (1999).

I agree that the traditional idea of free will may appear utterly mysterious and obscure in modern times unless we learn to think about free will in entirely new ways, to think in new directions, so to speak; and that is what I have been attempting to do in my work over four decades. Like many another issue of modernity, the question is whether something of the traditional idea of free will “in the superlative metaphysical sense” can be retrieved from the dissolving acids of modern science and secular learning or whether free will in the traditional sense will become, along with other aspects of our self-image, yet another example of the “disenchantments” of modernity.

## 2. Surface and Deeper Freedom and Ultimate Responsibility

The first question to address is why this traditional idea of free will was thought to be incompatible with necessity or determinism? We can begin to see why it might be thought so by reflecting on two familiar notions we understand—or think we understand—freedom and responsibility.

Nothing could be more important than freedom to the modern age. People clamor for it all over the world, often against authoritarian and violent resistance. And why do they want it? The simple, and not totally adequate, answer is that to be free is to be able to satisfy one's desires or do whatever one wants. In free societies, people can buy what they want, travel where they please, choose what to read, and so on. But these freedoms are what you might call *surface* freedoms. What is meant by *free will* runs deeper than these ordinary freedoms.

To see how, suppose we had maximal freedom to make choices of the above kinds to satisfy our desires, yet the choices we actually made were in fact manipulated by others, by the powers that be. In such a world we would have a great deal of everyday freedom to do whatever we wanted, yet our freedom of *will* would be severely limited. We would be free to *act* or to choose *what we willed*, but we would not have the ultimate power over what it is that we willed. Other persons would be pulling the strings, not by coercing or forcing us to do things against our wishes, but by manipulating us into having the wishes they wanted us to have. One sign of how important free will is to us is that people feel revulsion at such manipulation and feel demeaned by it when they find out it has been done to them. When subjected to it, they realize they were not their own persons; and having free will is about being your own person.

The centrality of this problem for modernity is illustrated by the popularity of 20<sup>th</sup> century dystopian works, such as Huxley's *Brave New World* or Skinner's *Walden Two*, and many other more recent incarnations in novels and films. In the futuristic societies described in these influential works, people can have and do whatever they will

or choose, but only to the extent that they have been conditioned since birth by behavioral engineers or neuro-chemists to will or choose only what they can have and do. Their surface freedoms are bought at the expense of a deeper freedom of the will.

Skinner goes further in a modern vein by arguing that this so-called deeper “freedom of the will” is no loss at all, since it is not something we can have anyway. In our ordinary lives, we are just as much the products of upbringing and social conditioning as the citizens of Walden Two, though we may delude ourselves into thinking otherwise. We may think we are the creators or originators of our own wills only because we are unaware of most of the genetic, psychological and social influences upon us. Then, echoing Nietzsche, Skinner adds that the idea that we could be ultimate or “original” creators of our own wills—that we could somehow be “causes of ourselves”—is an impossible ideal in any case, dreamt up by philosophers and theologians before we understood more about the hidden causes of behavior. It is an outdated idea that has no place in the modern scientific picture of the world.

Reflecting in this way on the idea of *freedom* is one path to understanding free will. Another is by reflecting on the notion of *responsibility*. Suppose a young man is on trial for an assault and robbery in which his victim was beaten to death. Let us say we attend his trial and listen to the evidence in the courtroom. At first, our thoughts of the young man are filled with anger and resentment. What he did was horrible. But as we listen daily to how he came to have the mean character and perverse motives he did have—a sad story of parental neglect, child abuse, sexual abuse, bad role models—some of our resentment against the young man is shifted over to the parents and others who abused and mistreated him. We begin to feel angry with them as well as with him. Yet we aren’t quite ready to shift all of the blame away from the young man himself. We wonder whether some residual responsibility may not belong to him. Our questions become: To what extent is *he* responsible for becoming the sort of person he now is? Was it *all* a question of bad parenting, societal neglect, social conditioning, and the like, or did he have any role to play in it?

These are crucial questions about free will and they are questions about what may be called the young man’s *ultimate responsibility*. We know that parenting and society, genetic make-up and upbringing, have an influence on what we become and what we are. But were these influences entirely *determining* or did they “leave anything over” for us to be responsible for? That is what we want to know about the young man. The question of whether he is merely a victim of bad circumstances or has some residual responsibility for being what he is—the question, that is, of whether he became the person he is *of his own free will*—seems to depend on whether these other factors were or were not *entirely* determining.

Reflections such as these point to a basic condition that throughout history has fueled intuitions that free will and determinism must be incompatible. I call it the condition of *ultimate responsibility* or UR, for short. The basic idea is this: to be ultimately responsible for an action, an agent must be responsible for anything that is a sufficient reason (condition, cause or motive) for the action's occurring.<sup>2</sup> If, for example, a choice issues from, and can be sufficiently explained by, an agent's character and motives (together with background conditions), 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.

This condition of ultimate responsibility or UR does not require that we could have done otherwise for *every* act done "of our own free wills." But it *does* require that we could have done otherwise with respect to *some* acts in our past life histories by which we formed our present characters. I call these "self-forming actions," or SFAs. 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 actions or SFAs) for which we could have done otherwise. If this were not so, there is *nothing we could have ever done 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 what we are.

Focusing on this condition of ultimate responsibility or UR tells us something else of great importance. It also tells us why the free will issue is about the freedom *of the will* and not just about the freedom of action. There has been a tendency in the modern era, beginning with Hobbes and Locke in the 17<sup>th</sup> century, to reduce the problem of free will to a problem of free action. I have been arguing for some time that such a reduction oversimplifies the problem.<sup>3</sup> Free will is not just about free action. It is about *self-formation*, about the formation of our "wills" or how we got to be the kinds of persons we are, with the characters, motives and purposes we now have. Were *we* ultimately responsible to some degree for having the wills we do have, or can the sources of our wills be completely traced backwards to something over which we had no control, such as Fate or the decrees of God, or heredity and environment or social conditioning or

---

2 For a formal statement and defense of this condition, see Kane (1996 ch. 3).

3 See the works listed in note 1.

hidden controllers, and so on? Therein, I believe, lies the core of the traditional problem of “free will.”

### 3. The Dialectic of Selfhood

Focusing on UR also shows something important about how free will is related to selfhood. I explained this in an earlier work in terms of what I called a “dialectic of selfhood.”<sup>4</sup> (The triadic structure will remind one of Hegel, but the details of this dialectic are mine.) In the first stage of this dialectic, imagine a baby several months old lying in a crib or infant seat. The baby’s arms and legs shake with uncontrolled and undirected energy as she looks about the room. This shaking comes from her nervous system, and ultimately from the brain which soaks up a high percentage of the energy-producing glucose of the body. (We call children “bundles of energy” for a reason.) The baby doesn’t know what to do with all that energy yet; her task is to gradually learn to get more control over it.

An early stage of this process of gaining control is one many parents have observed. Objects pass in front of the infant and she follows them with her eyes. She has no control over most of the objects and simply observes them pass by. But one passing object has a special fascination —her own hand. It is different, for it seems she can control it. One day she actually learns to hold the hand still in her visual field, make a fist with it, and then open it again. This turns out to be utterly fascinating. When she first discovers it, the act is repeated over and over again, and she smiles with delight at her success. She has discovered that this passing object is something special. It is part of her; and she can control it by an act of will. She has discovered the phenomena of *action* and *will* simultaneously by recognizing that she can control and direct some things out there in the world by attending to them and willing them to happen in her mind. No wonder she is fascinated.

Not surprisingly, this discovery is also connected to the distinction the infant is learning to make between herself and the world. And she begins to make this distinction in terms of what she can directly control with her will and what she cannot. Our full sense of *being a distinct self is tied up with our conception of being a distinct source of motion or activity in the world*, such that what goes on behind the screen of our mind (our will) can have effects out there in the world.

But in the second stage of the dialectic of selfhood, doubts arise about this simple picture. For we find that we are not separate from the world, but in it, and influenced by it in many hidden ways. Behind the window to the world —where *we* are supposed to be— is the brain, which is a physical object, like the body itself, part of world

---

4 See Kane (1996 91-97) for a fuller development of this dialectic.

and influenced by it. Perhaps we only *seem* to “move ourselves” by our wills in a primordial way, when we are in fact moved by causes coming from the world of which we are unaware operating through our brains and bodies. Such thoughts provoke a spiritual crisis. One crude reaction is to insist that we are not in the natural world at all—that the self behind the window is outside the natural world altogether, yet able to influence what goes on in that world in some magical way. A more subtle reaction is to argue that, while the world influences us, we can determine just *how* the world influences us through our senses and through our processing of information. We can determine what gets in and what is screened out, what influences our thought and action and what does not.

Alas, this solution only temporarily quells doubts about the influence of the world upon us. If we have already learned we are influenced by many things of which we are unaware, how can we be sure the very selections we make from within our inner sanctum are not determined by influences from the world in our past and present of which we are unaware and are beyond our control? What if our choices about *how* the world will influence us are *themselves* determined by the world? This thought propels us to a third stage of the dialectic of selfhood, where we encounter full-fledged threats of deterministic doctrines in all their historical guises—physical, biological, psychological, social, and so on. What I am suggesting is that we view the problem of determinism and free will, not as an isolated problem, but as a *stage in the dialectic of selfhood*—the process of self-understanding about the relation of our self with the world. At each stage, we are trying to preserve a remnant of the idea that we are in some sense *ultimately responsible* to some degree for how the world influences us and how we react to it—against the threat that we are merely products of forces coming wholly from the world.

Finally, and most importantly for our purposes, focusing on this condition of ultimate responsibility or UR shows us why free will has been historically thought to be incompatible with determinism. If agents must be responsible to some degree for anything (such as their prior formed character) that is a sufficient cause or motive for their actions, an impossible infinite regress of past actions would be required unless some actions in an agent’s life history did not have either sufficient causes or motives (and hence were not entirely determined). These undetermined actions would be the self-forming actions or SFAS, mentioned earlier, that are required by UR for self-formation and free will.

#### 4. The Intelligibility Question

But this approach to the incompatibility of free will and determinism through UR raises a host of further extremely difficult questions about free will—including how actions lacking

both sufficient causes *and* motives could themselves be free and responsible actions, and how, if at all, such actions could exist in the natural order where we humans live and have our being. These are versions of what I call the Intelligibility Question about free will, to which I now turn. Can we make sense of such a notion of free will or is it an unintelligible, impossible or self-contradictory ideal, as Nietzsche, Skinner and many other modern philosophers and scientists contend? And can such a notion of free will be reconciled with modern scientific conceptions of humans and the cosmos?

Doubts about the very possibility or intelligibility of free will are connected to an ancient dilemma: If free will is not compatible with *determinism*, it does not seem to be compatible with *indeterminism* either. Determinism means: Given the past, there is only one possible future. Indeterminism means the opposite: Same past, different possible futures. (Indeterminism suggests a “garden of forking paths” into the future in the image of Borges’s well-known story.) But how is it possible, one might ask, that different actions could arise voluntarily and intentionally from (exactly) the same past without occurring merely by luck or chance? This question has had a hypnotic effect on those who think about free will. One imagines that if free choices are undetermined, then which one occurs must be like spinning a wheel in one’s mind or one must just pop out by chance or randomly. If, for example, a choice occurred as a result of a quantum jump or other undetermined event in one’s brain, would that amount to a free and responsible choice? I’ll not trouble you with all the arguments, like these and others, by which philosophers have made the case that if choices or actions really were undetermined, then such choices or actions would occur as a matter of chance and hence would be “arbitrary,” or “capricious,” or “random,” “irrational,” “inexplicable,” mere matters of “luck” and not under the “control” of the agents, hence not free and responsible actions at all.

No wonder “libertarians” about free will —those who believe in a free will that is incompatible with determinism— have looked for some *deus ex machina* or other to solve the problem, while their opponents have cried magic or mystery. Indeterminism was required for free will, libertarians argued, but indeterminism was not enough. Indeterminism might provide causal gaps in nature. But that was only a negative condition. Some additional form of agency or causation was needed that went beyond causation in the natural order, whether deterministic or indeterministic. Thus, in response to modern science, we had numerous historical appeals in modernity, from Descartes to Kant and beyond, to “extra factors” such as noumenal selves, immaterial minds, uncaused causes, transempirical power centers, non-event agent causes, and the like, to account for a traditional libertarian free will. I long ago became



disenchanted with all such appeals and set myself the task of trying to find entirely new ways of thinking about free will that would not require such appeals.

### 5. Indeterminism and Responsibility

If one is to make sense of free will in a modern context, I believe one must avoid all such traditional strategies and take a whole new look at the indeterminist problem from the ground up. It is a scientific question, of course, whether the indeterminism *is* there in nature in appropriate ways. As the Epicureans said, if the atoms don't "swerve" in undetermined ways there would be no room in nature for free will. But our question is the philosophical one that has boggled people's minds for centuries: What could we *do* with indeterminism, assuming it *was* there in nature, to make sense of free will as something other than *mere* chance or randomness? Chance after all is not freedom. The first step in addressing this question is to note that indeterminism does not have to be involved in all acts done "of our own free wills" for which we are ultimately responsible, as argued earlier. Not all such acts have to be undetermined, but only those by which we made ourselves into the kinds of persons we are, namely "self-forming actions" or SFAS.

Now I believe these undetermined self-forming actions or SFAS occur at those difficult times of life when we are torn between competing visions of what we should do or become. Perhaps we are torn between doing the moral thing or acting from ambition, or between powerful present desires and long term goals, or we are faced with difficult tasks for which we have aversions. In all such cases, we are faced with competing motivations and have to make an effort to overcome temptation to do something else we also strongly want. There is tension and uncertainty in our minds about what to do at such times, I suggest, that is reflected in appropriate regions of our brains by movement away from thermodynamic equilibrium—in short, a kind of "stirring up of chaos" in the brain that makes it sensitive to micro-indeterminacies at the neuronal level. The uncertainty and inner tension we feel at such soul-searching moments of self-formation is thus reflected in the indeterminacy of our neural processes themselves. What is experienced internally as uncertainty would then correspond physically to the opening of a window of opportunity that would temporarily screen off complete determination by influences of the past.

When we do decide under such conditions of uncertainty, the outcome is not determined because of the preceding indeterminacy—and yet it can be willed (and hence rational and voluntary) either way owing to the fact that in such self-formation, the agents' prior wills are divided by conflicting motives. Consider a businesswoman who faces such a conflict. She is on her way to an important

meeting when she observes an assault taking place in an alley. An inner struggle ensues between her conscience, to stop and call for help, and her career ambitions which tell her she cannot miss this meeting. She has to make an effort of will to overcome the temptation to go on. If she overcomes this temptation, it will be the result of her effort, but if she fails, it will be because she did not *allow* her effort to succeed. And this is due to the fact that, while she willed to overcome temptation, she also willed to fail, for quite different and incommensurable reasons. When we, like the woman, decide in such circumstances, and the indeterminate efforts we are making become determinate choices, we *make* one set of competing reasons or motives prevail over the others then and there *by deciding*.

Now let us add a further piece to the puzzle. Just as indeterminism need not undermine rationality and voluntariness, so indeterminism in and of itself need not undermine control and responsibility. Suppose you are trying to think through a difficult problem, say a mathematical problem, and there is some indeterminacy in your neural processes complicating the task—a kind of chaotic background. It would be like trying to concentrate and solve a problem, say a mathematical problem, with background noise or distraction. Whether you are going to succeed in solving the problem is uncertain and undetermined because of the distracting neural noise. Yet, if you concentrate and solve the problem nonetheless, there is reason to say you did it and are responsible for it even though it was undetermined whether you would succeed. The indeterministic noise would have been an obstacle that you overcame by your effort.

There are numerous examples supporting this point, where indeterminism functions as an obstacle to success without precluding responsibility. Consider an assassin who is trying to shoot the prime minister, but might miss because of some undetermined events in his nervous system that may lead to a jerking or wavering of his arm. If the assassin does succeed in hitting his target, despite the indeterminism, can he be held responsible? The answer is clearly yes because he intentionally and voluntarily succeeded in doing what he was *trying* to do—kill the prime minister. Yet his action, killing the prime minister, was undetermined. It might have failed. Or, here is another example: a husband, while arguing with his wife, in a fit of rage swings his arm down on her favorite glass-top table top intending to break it. Again, we suppose that some indeterminism in his outgoing neural pathways makes the momentum of his arm indeterminate, so that it is genuinely undetermined whether the table will break right up to the moment when it is struck. Whether the husband breaks the table or not is undetermined and yet he is clearly responsible if he does break it. (It would be a poor excuse for him

to say to his wife: “chance did it, not me.” Even though there was a chance he wouldn’t break it, chance didn’t do it, *he* did.)

Now these examples — of the mathematical problem, the assassin and the husband— are not all we want, since they do not amount to genuine exercises of (self-forming) free will in SFAS, like the businesswoman’s, where the will is divided between conflicting motives. The assassin’s will is not divided between conflicting motives as is the woman’s. He wants to kill the prime minister, but does not also want to fail. (If he fails therefore, it will be *merely* by chance.) Yet these examples of the assassin, the husband and the like, do provide some clues. To go further, we have to add some further thoughts.

Imagine in cases of inner conflict characteristic of SFAS, like the businesswoman’s, that the indeterministic noise which is providing an obstacle to her overcoming temptation is not coming from an external source, but is coming from her own will, since she also deeply desires to do the opposite. Imagine that two crossing (recurrent) neural networks are involved, each influencing the other, and representing her conflicting motivations. (Recurrent neural networks are complex networks of interconnected neurons in the brain circulating impulses in feedback loops that are now generally thought to be involved in higher-level cognitive processing.)<sup>5</sup> The input of one of these neural networks consists in the woman’s reasons for acting morally and stopping to help the victim; the input of the other, her ambitious motives for going on to her meeting.

The two networks are connected so that the indeterministic noise which is an obstacle to her making one of the choices is coming from her desire to make the other, and vice versa — the indeterminism thus arising from a tension-creating conflict in the will, as I said. In these circumstances, when either of the pathways reaches an activation threshold (which amounts to choice), it will be like your solving the mathematical problem by overcoming the background noise produced by the other. And just as when you solved the mathematical problem by overcoming the distracting noise, one can say you did it and are responsible for it, so one can say this as well, I argue, in the present case, *whichever one is chosen*. The pathway through which the woman succeeds in reaching a choice threshold will have overcome the obstacle in the form of indeterministic noise generated by the other.

Note that, under such conditions, the choices either way will not be “inadvertent,” “accidental,” “capricious,” or “merely random,” (as critics of indeterminism say) because they will be *willed* by the agents either way when they are made, and done for *reasons* either

---

5 Readable and accessible introductions to the role of neural networks (including recurrent networks) in cognitive processing include Churchland (1996) and Spitzer (1999).

way —reasons that the agents then and there *endorse*. But these are the conditions usually required to say something is done “on purpose,” rather than accidentally, capriciously or merely by chance. Moreover, these conditions taken together, I have argued, rule out each of the reasons we have for saying that agents act, but do not have *control* over their actions (compulsion, coercion, constraint, inadvertence, accident, control by others, etc.).<sup>6</sup>

Indeed, in these cases, agents have what I call “plural voluntary control” over the options in the following sense: They are able to bring about *whichever* of the options they will, *when* they will to do so, for the *reasons* they will to do so, on *purpose* rather than accidentally or by mistake, without being coerced or compelled in doing so or willing to do so, or otherwise controlled in doing or willing to do so by any other agents or mechanisms. I show in my 1996 book and elsewhere that each of these conditions can be satisfied for SFAS as conceived above even though the SFAS are undetermined (e.g. Kane 1996, chapter 8). The conditions can be summed up by saying, as we sometimes do, that the agents can choose either way, *at will*.

Note also that this account of self-forming choices amounts to a kind of “doubling” of the mathematical problem. It is as if an agent faced with such a choice is *trying* or making an effort to solve *two* cognitive problems at once, or to complete two competing (deliberative) tasks at once —in our example, to make a moral choice and to make a conflicting self-interested choice (corresponding to the two competing neural networks involved). Each task is being thwarted by the indeterminism coming from the other, so it might fail. But if it succeeds, then the agents can be held responsible because, as in the case of solving the mathematical problem, they will have succeeded in doing what they were knowingly and willingly trying to do. Recall the assassin and the husband. Due to indeterminacies in their neural pathways, the assassin might miss his target or the husband fail to break the table. But if they *succeed*, despite the probability of failure, they are responsible, because they will have succeeded in doing what they were trying to do.

And so it is, I suggest, with self-forming choices or SFAS, except that in the case of self-forming choices, *whichever way the agents choose* they will have succeeded in doing what they were trying to do

---

6 We have to make further assumptions about the case in order to rule out some of these conditions. For example, we have to assume, no one is holding a gun to the woman’s head forcing her to go back, or that she is not paralyzed, etc. But the point is that none of these conditions is inconsistent with the case of the woman as we have imagined it. If these other conditions are satisfied, as they can be, and the businesswoman’s case is in other respects as I have described it, we have an SFA. For the complete argument see Kane (1996 ch. 8), among other works listed in note 1.

because they were simultaneously trying to make both choices, and one is going to succeed. Their failure to do one thing is not a *mere* failure, but a voluntary succeeding in doing the other.

Does it make sense to talk about the agent's trying to do two competing things at once in this way, or to solve two cognitive problems at once? Well, much current scientific evidence points to the fact that the brain is a parallel processor; it simultaneously processes different kinds of information relevant to tasks such as perception or recognition through different neural pathways. Such a capacity, I believe, is essential to the exercise of free will. In cases of self-formation (SFAS), agents are simultaneously trying to resolve plural and competing cognitive tasks. They are, as we say, of two minds. Yet they are not two separate persons. They are not dissociated from either task. The businesswoman who wants to go back to help the victim is the same ambitious woman who wants to go to her meeting. She is 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 free will. And when she succeeds in doing one of the things she is trying to do, she will endorse that as *her* resolution of the conflict in her will, voluntarily and intentionally, not by accident or mistake.

## 6. Responsibility, Luck and Chance

Now you may find this interesting and yet still find it hard to shake the intuition that if choices are undetermined, they *must* happen merely by chance —and so must be “random,” “capricious,” “uncontrolled,” “irrational,” and all the other things charged. Such intuitions are deeply ingrained. But if we are ever going to understand free will, I think will have to break old habits of thought that support such intuitions. The first step is to question the intuitive connection in most people's minds between “indeterminism's being involved in something” and “its happening merely as a matter of chance or luck.” “Chance” and “luck” are terms of ordinary language that carry the connotation of “its being out of my control.” So using them already begs certain questions, whereas “indeterminism” is a technical term that merely precludes *deterministic* causation, though not causation altogether. Indeterminism is consistent with nondeterministic or probabilistic causation, where the outcome is not inevitable. It is therefore a mistake (alas, one of the oldest and most common in debates about free will) to assume that “undetermined” means “uncaused.”

Here is another source of misunderstanding. Since the outcome of the businesswoman's effort (the choice) is undetermined up to the last minute, we may have the image of her first making an effort to overcome the temptation to go on to her meeting and then at the last instant “chance takes over” and decides the issue for her. But

this is misleading. One cannot separate the indeterminism and the effort of will, so that *first* the effort occurs *followed* by chance or luck (or vice versa). The effort *is* indeterminate and the indeterminism is a property of the effort, not something separate that occurs after or before the effort. The fact that the effort has this property of being indeterminate does not make it any less the woman's *effort*. The complex recurrent neural network that realizes the effort in the brain is circulating impulses in feedback loops and there is some indeterminacy in these circulating impulses. But the whole process is her effort of will and it persists right up to the moment when the choice is made. There is no point at which the effort stops and chance "takes over." She chooses as a result of the effort, even though she might have failed. Similarly, the husband breaks the table as a result of his effort, even though he might have failed because of the indeterminacy. (That is why his excuse, "chance broke the table, not me" is so lame.)

Just as expressions like "she chose *by chance*" can mislead in such contexts, so can expressions like "she got lucky." Recall that, with the assassin and husband, one might say "they got lucky" in killing the prime minister and breaking the table because their actions were undetermined. *Yet they were responsible*. So ask yourself this question: why does the inference "he got lucky, so he was *not* responsible?" fail in the cases of the husband and the assassin where it does fail? The first part of an answer has to do with the point made earlier that "luck," like "chance," has question-begging implications in ordinary language that are not necessarily implications of "indeterminism" (which implies only the absence of deterministic causation). The core meaning of "he got lucky" in the assassin and husband cases, which is implied by indeterminism, I suggest, is that "he succeeded *despite the probability or chance of failure*"; and this core meaning does not imply lack of responsibility, *if he succeeds*.

If "he got lucky" had other meanings in these cases that are often associated with "luck" and "chance" in ordinary usage (for example, the outcome was not his doing, or occurred by *mere chance*, or he was not responsible for it), the inference would not fail for the husband and assassin, as it clearly does. But the point is that these further meanings of "luck" and "chance" do not follow *from the mere presence of indeterminism*. The second reason why the inference "he got lucky, so he was not responsible" fails for the assassin and the husband is that *what* they succeeded in doing was what they were *trying* and wanting to do all along (kill the minister and break the table respectively). The third reason is that *when* they succeeded, their reaction was not "oh dear, that was a mistake, an accident —something that *happened* to me, not something I *did*." Rather they *endorsed* the outcomes as something they were trying and wanting to do all along, that is to say, knowingly and purposefully, not by mistake or accident.

But these conditions are satisfied in the businesswoman's case as well, *either way* she chooses. If she succeeds in choosing to return to help the victim (or in choosing to go on to her meeting) (i) she will have "succeeded *despite the probability or chance of failure*," (ii) she will have succeeded in doing what she was *trying* and *wanting* to do all along (she wanted both outcomes very much, but for different reasons, and was trying to make those reasons prevail in both cases), and (iii) when she succeeded (in choosing to return to help) her reaction was not "oh dear, that was a mistake, an accident —something that happened to me, not something I did." Rather she *endorsed* the outcome as something she was trying and wanting to do all along; she recognized it as her resolution of the conflict in her will. And if she had chosen to go on to her meeting she would have endorsed that outcome, recognizing it as her resolution of the conflict in her will.

Well, if indeterminism does not undermine the idea that the woman's choices are purposeful, does it undermine the idea that it is the *woman*, the *agent*, who makes the choices? The answer again is no. Indeterminism is consistent with *agency*, as we have seen in the cases of the assassin and the husband, *when it is an ingredient in some larger goal-directed or teleological process*; and that is how I envisage the efforts of will leading to self-forming choices. To explain these larger goal-directed processes in modern terms, I would argue that we must appeal to "*dynamical systems theory*" or the theory of "complex dynamical systems," in which the rational agent is viewed as "a hierarchically ordered, information sensitive, complex dynamical system."

In dynamical systems of these sorts generally, which are now widely recognized in nature, complex wholes can causally influence their parts in a top-down manner consistent with the usual bottom-up causation of physical parts to wholes. The first important steps toward applying dynamical systems theory to human action have been taken in a number of recent works, including Alicia Juarrero's *Dynamics in Action* and Nancey Murphy and Warren Brown's *Did My Neurons Make Me Do It?*, among other works. I agree with the broad outlines of the account of human agency presented in these and other works. That account of agency must be expanded, of course, to give a full account of free will as I have been describing it, adding other elements, including the element of indeterminism. But I believe these additions, including indeterminism, are consistent with the broader outlines of the dynamical systems approach and human agency generally, as I have been arguing. And I would add an historical/philosophical note: The idea of rational agents as "hierarchically ordered, information sensitive, complex dynamical systems" is very much in the tradition of Aristotle's "form/matter" account of the human agent, a version of which I would endorse.

### 7. Control, Indeterminism, Efforts and Introspection

But would not the presence of indeterminism, if not entirely undermine responsible agency, at least *diminish* the control persons have over their choices and other actions? Is it not the case that the assassin's control over whether the prime minister 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? And this limitation seems to be connected with another problem often noted by critics of incompatibilist freedom —the problem that indeterminism, wherever it occurs, seems to be a *hindrance* or *obstacle* to our realizing our purposes and hence an obstacle to (rather than an enhancement of) our freedom.

There is a truth to these claims, but I think what is true in them also reveals something important about free will. 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 that in the case of the businesswoman (and SFAS generally), the indeterminism that is admittedly diminishing her control over one thing she is trying to do (the moral act of helping the victim) *is coming from her own will* —from her desire and effort to do the opposite (go to her business meeting). And the indeterminism that is diminishing her control over the other thing she is trying to do (act selfishly and go to her meeting) is coming from her desire and effort to do the opposite (to be a moral person and act on moral reasons). So, 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 would indeed in a sense have “complete control” over one of her options. There would be no competing motives that would stand in the way of her choosing it. But then also she would not be free to rationally and voluntarily choose the other purpose 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 paradoxically opens up the genuine possibility of pursuing other purposes —of choosing or doing *otherwise* in accordance with, rather than against, our wills (voluntarily) and reasons (rationally). 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. Self-formation is not a gift, but a struggle.

Recall Kant's image of the bird which is upset by the resistance of the air and the wind to its flight and so imagines that it could fly better if there were no air at all to resist it. But, of course, as Kant notes, the bird would not fly better if there were no wind, but would



cease to fly at all. So it is with indeterminism and free will. It provides resistance to our choices, but a resistance that is necessary if we are to be self-forming agents. And being such self-forming agents is deeply connected to our sense of self, as I have argued.

Another concern that has been raised about the above account of free will is that we are not introspectively aware of making dual efforts and performing multiple cognitive tasks in such choice situations. But I am not claiming that agents are conscious of making dual efforts. What they are introspectively conscious of is that they are trying to decide about which of two options to choose and that either choice is a difficult one because there are resistant motives pulling them in different directions that will have to be overcome, whichever choice is made. In such introspective conditions, I am theorizing that what is actually going on underneath is a kind of parallel processing in the brain that involves separate efforts or endeavors to resolve competing cognitive tasks. The point is that introspective evidence does not give us the whole story about free will. If we stay on the surface and just consider what our immediate experience tells us, free will is bound to appear mysterious, as it has appeared to so many people through the centuries. To unravel its mysteries, we have to consider what might be going on behind the scenes.

As noted earlier, it is now widely believed that parallel processing takes place in the brain in such cognitive phenomena as visual perception. The theory is that the brain separately processes different features of the visual scene, such as object and background, through distinguishable and parallel, though interconnected, neural pathways or “streams.”<sup>7</sup> Suppose someone objected that we are not introspectively aware of such distributed processing in ordinary cases of perception. That would hardly be a decisive objection to this new theory of vision. For the claim is that this is what we are doing in visual perception, not necessarily that we are introspectively aware of doing it. And I am making a similar claim about free will. If parallel processing is involved in the *input* side of the cognitive ledger (in perception), then why not consider that it might be involved in the *output* side as well (in deliberation and choice)? What is needed is a *theory* about what might be going on when we exercise free will, not merely a description of what we immediately experience.

It has also been objected that it is irrational to make efforts to do incompatible things. I concede that in most ordinary situations it is. But I argue that there are special circumstances in which it is not irrational to make competing efforts: These include circumstances in which (i) we are deliberating between competing options; (ii) we intend to choose one or the other, but cannot choose both; (iii) we have powerful motives for wanting to choose each of the options for

---

7 For further discussion, see Bechtel et. al. (2001).

different and incommensurable reasons; (iv) there is a consequent resistance in our will to either choice, so that (v) if either choice is to have a chance of being made, effort will have to be made to overcome the temptation to make the other choice; and most importantly, (vi) we want to give each choice a fighting chance of being made because the motives for each choice are important to us. The motives for each choice define in part what sort of person we are; and we would be taking them lightly if we did not make an effort on their behalf. These conditions are, of course, the conditions of SFAS.

### 8. *Liberum Arbitrium Voluntatis*

I conclude with one final objection. Even if one granted that persons, such as the businesswoman, could make genuine self-forming choices that were undetermined, isn't their something to the charge that such choices would be *arbitrary*? A residual arbitrariness seems to remain in all self-forming choices since the agents cannot in principle have sufficient or overriding *prior* reasons for making one option and one set of reasons prevail over the other. This is again one of those truths that tells us something important about free will. In this case, as I have discussed elsewhere, it tells us that every undetermined self-forming free choice is the initiation of what I have called a "value experiment" whose justification lies in the future and is not fully explained by past reasons. In making such a choice we say, in effect, "Let's try this. It is not required by my past, but it is consistent with my past and is one branching pathway my life can now meaningfully take. Whether it is the right choice, only time will tell. Meanwhile, I am willing to take responsibility for it one way or the other" (Kane 1996 145-6).

The term "arbitrary" as I have often noted comes from the Latin *arbitrium*, which means "judgment" —as in *liberum arbitrium voluntatis*, "free judgment of the will" (the medieval philosophers' 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 author's of, and characters in, their own stories all 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 past that, if they are truly free, does not limit their future pathways to one. In response to the charge that they did not have sufficient or *conclusive* prior reasons for choosing as they did, they may respond. "True

enough. But I did have *good* reasons for choosing as I did, which I'm willing to stand by *and take responsibility for*. If they were not sufficient or conclusive reasons, that's because, like the heroine of the novel, I was not a fully formed person before I chose (and still am not, for that matter). 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." To be both author and character in one's own story all at once is what it means I believe to be a *self*.

### **Bibliography**

- Bechtel, W. et. al., eds. *Philosophy and the Neurosciences: A Reader*. Malden: Blackwell, 2001.
- Churchland, P. M. *The Engine of Reason, the Seat of the Soul*. Cambridge: MIT Press, 1996.
- Huxley, A. *Brave New World*. New York: Harper-Collins, 1989.
- Juarrero, A. *Dynamics in Action*. Cambridge: MIT Press, 1999.
- Kane, R. *Free Will and Values*. New York: State University of New York Press, 1985.
- Kane, R. *The Significance of Free Will*. New York: Oxford University Press, 1996.
- Kane, R. "Responsibility, Luck and Chance: Reflections on Free Will and Indeterminism", *Journal of Philosophy* 96 (1999): 217-40.
- Kane, R., ed. *The Oxford Handbook of Free Will*. New York: Oxford University Press, 2002a.
- Kane, R., ed. *Free Will*. Malden: Blackwell, 2002b.
- Kane, R. *A Contemporary Introduction to Free Will*. New York: Oxford University Press, 2005.
- Murphy, N. & Brown, W. *Did My Neurons Make Me Do It?* New York: Oxford University Press, 2007.
- Nietzsche, F. *Beyond Good and Evil*, Norman, J., trad. Cambridge: Cambridge University Press, 2002.
- Skinner, B. F. *Walden Two*. London: MacMillan, 1962.
- Spitzer, M. *The Mind Within the Net*. Cambridge: MIT Press, 1999.