

---

**RESPONSIBILITY AND FREE WILL IN DWORKIN'S  
*JUSTICE FOR HEDGEHOGS***

**ROBERT KANE\***

There is much I agree with in Ronald Dworkin's impressive work, *Justice for Hedgehogs*, concerning objective value, living well, respect, and many other things. On all these matters, it is a work that deserves to be widely read and pondered. Specifically, I agree with much of Dworkin's Chapter Ten, entitled "Free Will and Responsibility," which is the specific subject of this paper. Perhaps ninety percent of what he says about free will and responsibility in that chapter I agree with. But, following the perverse custom of philosophers, and because important issues are at stake, I will focus here on the ten percent disagreement. My disagreements stem from forty years of dealing with the philosophical problem of free will and from the fact that, unlike Dworkin, I am an incompatibilist about free will and moral responsibility. I believe neither is compatible with a thoroughgoing determinism of human behavior.

Dworkin offers two opposing models to explain the freedom required for judgmental responsibility: the "hydraulic" model – according to which an act is yours and you are responsible for it, if it has its origin in an uncaused act of will – and the "creative" model – according to which an act is yours and you are responsible for it if you have the capacities (i) to form pertinent beliefs about the world, and (ii) to match your decisions to your normative personality.<sup>1</sup> The first, the hydraulic model, he argues, is *incompatible* with determinism, but it is deeply flawed.<sup>2</sup> Uncaused acts of will are mysterious, an outdated remnant from earlier ages, and have no place in the modern scientific picture of the world. The creative model, by contrast, captures what we mean by freedom and responsibility in everyday contexts, and it is *compatible* with determinism.<sup>3</sup>

Now, I agree completely that what he calls the hydraulic model is a non-starter, an outdated remnant of earlier times.<sup>4</sup> And I agree that the creative model captures much of what we require for freedom and judgmental

---

\* University Distinguished Teaching Professor of Philosophy, The University of Texas at Austin.

<sup>1</sup> RONALD DWORKIN, *JUSTICE FOR HEDGEHOGS* (forthcoming 2010) (Apr. 17, 2009 manuscript at 146-58, on file with the Boston University Law Review).

<sup>2</sup> *See id.* (manuscript at 147).

<sup>3</sup> *See id.* (manuscript at 154).

<sup>4</sup> *See id.* (manuscript at 152).

---

---

responsibility (my ninety percent agreement). But the dichotomy is too simple, suggesting that we must choose one or another of these two models, which is a mistake. One can reject the hydraulic model, as I do, and accept much of the creative model, as I would, and still believe that the traditional problem of free will is not really engaged, much less solved, by this choice. In my view, there has been a general tendency in the modern era, beginning with Hobbes and Locke through Hume and Mill and many other worthies, to oversimplify the traditional free will problem: to reduce it to a simpler problem and then to solve the simpler problem. As a result, free will in the traditional sense becomes yet another victim of the dissolving acids of modernity.

What is this traditional problem of free will that is thus deflated? I offer three clues, focusing throughout on what Dworkin calls moral responsibility,<sup>5</sup> where we can speak of blameworthiness for actions, for that is especially relevant to both moral and legal contexts as well as to the traditional problem of free will. The first clue comes from Aristotle, who said that if the circumstances in which you act, including the character, decisions, and motives from which you act, determine your action, then to be responsible for the action you must be responsible for at least some of the circumstances, such as your character, decisions, and motives, that led you to act as you did.<sup>6</sup> This statement is no esoteric principle. It is woven into the fabric of our everyday reasoning about responsibility. If a drunk driver kills a pedestrian, and it could be shown that given the circumstances, including the visibility on the road and the condition of his nervous system given the alcohol, he could not possibly have avoided hitting the pedestrian, that fact alone is not exonerating. One wants to know whether he is responsible for any of the prior circumstances that placed him in a position where he could not possibly have avoided hitting the pedestrian, such as his prior decisions to drink and then to drive, and so on.

The second clue involves a story. Imagine a young man 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 at his vile act. 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, and 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 are not quite ready to shift all of the blame away from the young man himself. We wonder whether some residual responsibility may not still belong to him. We ask ourselves, “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?”

---

<sup>5</sup> See, e.g., *id.* (manuscript at 7).

<sup>6</sup> See ARISTOTLE, THE NICOMACHEAN ETHICS bk. III, at 48-78 (J.L. Ackrill & J.O. Urmson eds., David Ross trans., Oxford Univ. Press 1992) (c. 384-322 B.C.E.).

---

---

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 the question we want answered 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. It is important that the young man be able to conform his behavior to his normative personality, as Dworkin's creative model requires.<sup>7</sup> But we also want to know whether and to what degree he is responsible for forming that "normative personality" to which he now conforms.

This story tells us something further that is also important about free will. An aspect of the reductive tendency of the modern era I mentioned earlier is to reduce the problem of free will to one of freedom of action. But freedom of will is not just about free action; it is about self-formation. How did we get to be the kinds of persons we are, with the wills (characters and motives) we now have? Was it *we* ourselves who were responsible for forming our characters and motives, or someone or something else beyond our control – God or Fate, heredity and environment, nature or upbringing, society or culture, hidden controllers, and so on? Therein, I believe, lies the traditional problem of free will.

The third clue to this traditional problem comes from the meaning of determinism itself. Determinism means that, given the past at any time and the laws of nature, there is only one possible future.<sup>8</sup> Indeterminism means the opposite: given the same past and laws, there are multiple possible futures. Indeterminism implies a forking paths view of the future, which is the way we normally think about freedom. Determinism, by contrast, is a very strong doctrine. It implies that, given the past and laws of nature, any act a person performs is the only act that could have been performed in the circumstances. So it would be for the young man. And the impossibility here is very strong, what the logicians formalize in the modal system S5: in no logically possible world whatever, can there be <this past, these laws, and any other act occurs but the one that actually did occur>.

It can be shown that this *is* the doctrine of determinism that has worried philosophers for centuries, whether in logical, theological, or scientific forms. Determinism is a strong and worrisome doctrine, as one can see by considering

---

<sup>7</sup> See DWORKIN, *supra* note 1 (manuscript at 155) ("Someone cannot lead a life if he has not formed a normative personality – a reasonably stable system of desires, preferences, tastes, convictions, attachments, loyalties, ideals, and the rest – and if he cannot make decisions that exhibit that personality.").

<sup>8</sup> See, e.g., *id.* (manuscript at 137).

what Israeli philosopher David Widerker has called the “W-defense.”<sup>9</sup> If persons are accused of having done something wrong they may ask, “What should I have done instead?” And suppose in the circumstances, there is nothing else they could have done. Is this fact exonerating? Not necessarily. For they may be responsible by virtue of prior actions for putting themselves in these circumstances where there is now nothing else they could have done. Thus, we may invoke the Aristotelian principle, mentioned earlier,<sup>10</sup> which as I said is woven into the fabric of our ordinary reasoning about responsibility. The problem is that, if determinism is true, there is nothing else they could have *ever* done differently in their entire lifetimes to make themselves *or* their circumstances any different than they are.

Compatibilists about free will and moral responsibility in the modern era, from Hobbes and Locke onward, have invoked a common strategy for avoiding this problem.<sup>11</sup> They have defended a hypothetical interpretation of “could have done otherwise”: all we mean when we say persons “could have done otherwise,” they argue, is that “they would have done otherwise, if the past had been different in some way,” if the persons had had different beliefs or desires, or reasoned or chosen differently, they would have acted differently.<sup>12</sup> But this strategy is deeply flawed. Immanuel Kant aptly called it a “wretched subterfuge”<sup>13</sup> and William James a “quagmire of evasion.”<sup>14</sup> Many objections can be made against this strategy, and even many prominent compatibilists today reject it. (To his credit, Dworkin does not appeal to this traditional strategy.) It may be true that persons would have done otherwise, if the past had been different in some way. But the past was not different in some way; it was as it was. Our freedom and responsibility must be exercised in the world that actually is, not in some hypothetical world that never was.

A second strategy, common to more recent compatibilists (to which Dworkin does appeal), is simply to deny that moral responsibility for actions requires that we could have done otherwise at all.<sup>15</sup> This strategy is more subtle, but also flawed, I believe, for more nuanced reasons. It is true that we

---

<sup>9</sup> David Widerker, *Frankfurt's Attack on the Principle of Alternative Possibilities: A Further Look*, in 14 *PHILOSOPHICAL PERSPECTIVES: ACTION AND FREEDOM* 181, 191 (James E. Tomberlin ed., 2000).

<sup>10</sup> See *supra* note 6 and accompanying text.

<sup>11</sup> See, e.g., THOMAS HOBBS, *LEVIATHAN* 136 (Edwin Curley ed., Hackett Publ'g Co. 1994) (1668).

<sup>12</sup> See *id.* (“[N]o liberty can be inferred to the will, desire, or inclination, but the liberty of the man; which consisteth in this: that he finds no stop, in doing what he has the will, desire, or inclination to do.”).

<sup>13</sup> IMMANUEL KANT, *Critical Examination of Practical Reason*, in *KANT'S CRITIQUE OF PRACTICAL REASON AND OTHER WORKS ON THE THEORY OF ETHICS* 85, 189 (Thomas Kingsmill Abbott trans., Longmans, Green & Co. 1948) (1788).

<sup>14</sup> WILLIAM JAMES, *The Dilemma of Determinism*, in *THE WILL TO BELIEVE AND OTHER ESSAYS IN POPULAR PHILOSOPHY* 145, 149 (Longmans, Green & Co. 1927) (1896).

<sup>15</sup> See DWORKIN, *supra* note 1 (manuscript at 154).

---

can be held responsible for many actions even though we could not have done otherwise *and* even if those actions were determined by our then-existing characters and motives. 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 choices and actions in the past 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 ultimately responsible for being the way we are. Thus there is a subtle fallacy of composition involved here: while it is true that we may be responsible for many actions, even when we could not have done otherwise, and even when the acts were determined, it does not follow that this could be true of *all* of our actions in an entire lifetime. Or, if you have karmic tendencies, it does not follow that this could be true of *all* of our actions in *all* of our lifetimes. Yet, determinism implies just that.

A third and final strategy of compatibilists, which is also employed by Dworkin, is to say that incompatibilist free will makes no sense because it reduces either to chance or to the mystery of uncaused causes which have no place in the modern scientific picture of the world.<sup>16</sup> Responding to this strategy is a much harder and longer task – one I have been working on for forty years and will only briefly discuss here, while directing readers to other writings.<sup>17</sup> I believe one can reconcile incompatibilist free will with modern science without supposing it reduces to mere chance or mystery. The task is not easy, but I believe many developments in the modern sciences themselves make it more likely than it may have seemed in the past.

One thing required is that we not confuse *causation* and *determination*, assuming that what is undetermined must thereby be uncaused (a mistake that is, alas, all too common in debates about free will). I agree with Dworkin that it is confused and mistaken to think that free choices, decisions, or acts must be entirely uncaused by prior events.<sup>18</sup> Our choices, decisions, and acts are causally influenced, as indicated earlier, by our heredity and environment, by social upbringing, by our desires, beliefs, motives, and by numerous other factors that influence our behavior. What is required is not that our free choices or acts be *uncaused*, but that they be *undetermined* – or, in other words, that the causal influences on them not always be entirely determining.

---

<sup>16</sup> See *id.* (manuscript at 146).

<sup>17</sup> The ideas that follow have been developed at greater length in a number of my previous works. See generally, e.g., ROBERT KANE, A CONTEMPORARY INTRODUCTION TO FREE WILL (2005); R. KANE, FREE WILL AND VALUES (1985); ROBERT KANE, THE SIGNIFICANCE OF FREE WILL (1996); Robert Kane, *Responsibility, Luck and Chance: Reflections on Free Will and Indeterminism*, 96 J. PHIL. 217 (1999) [hereinafter Kane, *Responsibility*]; Robert Kane, *Some Neglected Pathways in the Free Will Labyrinth*, in THE OXFORD HANDBOOK OF FREE WILL 406 (Robert Kane ed., 2002).

<sup>18</sup> DWORKIN, *supra* note 1 (manuscript at 152) (rejecting the hydraulic control principle as “popular as an interpretation of the more abstract idea that people are not judgmentally responsible for their acts when they are not in control,” yet ultimately “implausible”).

---

---

In addition to distinguishing causation from determination, other steps are necessary to make sense of a free will that involves indeterminism. One step is to recognize that, as neuroscientific evidence now tells us, the brain is a parallel processor.<sup>19</sup> It can simultaneously process different kinds of information relevant to tasks such as perception and recognition through different neural pathways. Such a capacity is essential, I believe, to the exercise of free will. In deliberation leading to “self-forming” choices, separate but interconnected cognitive processes aim at competing goals (different possible choices) represented by competing attractors in the brain. In such cases of self-formation, we are faced with competing motivations and are torn between competing visions of what we should do or become. The tension and uncertainty about what to do at such times, I suggest, is reflected in appropriate regions of our brains by movement away from thermodynamic equilibrium – a kind of “stirring up of chaos” in the brain that makes it sensitive to micro-indeterminacies at the neuronal level.<sup>20</sup> The tension and uncertainty we feel at such soul-searching moments would thus be reflected in the indeterminacy of our neural processes themselves.

When we do decide under such conditions of uncertainty, the outcome would not be determined because of the preceding indeterminacy; yet the outcome 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, for example, 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 (believing she should stop and call for help) and her career ambitions (telling her she cannot miss this meeting). She has to make an effort of will to overcome the temptation to go on to the meeting. If she overcomes this temptation, it will be the result of her effort. However, if she fails, it will be because she did not *allow* her effort to succeed. 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 businesswoman, 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*.

Just as indeterminism in such circumstances need not undermine rationality and voluntariness of choices, it also need not undermine action and responsibility. Suppose, for example, 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

---

<sup>19</sup> See, e.g., Alison B. Wismer Fries & Seth D. Pollack, *Emotion Processing and the Developing Brain*, in HUMAN BEHAVIOR, LEARNING, AND THE DEVELOPING BRAIN: TYPICAL DEVELOPMENT 329, 331 (Donna Coch et al. eds., 2007).

<sup>20</sup> Much of the subsequent discussion, including the examples that follow, are also discussed in Kane, *Responsibility*, *supra* note 17, at 224-27.

---

---

chaotic background noise impeding your concentration. Whether you are going to succeed in solving the problem is uncertain and undetermined because of the distracting neural noise. However, if you concentrate and solve the problem nonetheless, we have 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. Imagine an assassin who is trying to shoot a prime minister with a high-powered rifle, but who might miss because of some undetermined events in his nervous system that 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 – assassinate the prime minister. Yet his action, killing the prime minister, was undetermined.

Or consider a husband who, 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 undetermined whether the table will actually break until the moment when it is struck. Whether the husband breaks the table is undetermined and yet he is clearly responsible if he does break it. It would be a poor excuse to offer his wife if he claimed, “Chance did it, not me.” Though indeterminism was involved, chance did not do it, he did.

These are important clues, but further steps must be added. What we have to imagine in self-forming choices, such as the businesswoman’s, is that the indeterministic noise which is providing an obstacle to her overcoming temptation is not coming from an unwilled source (as in the assassin and husband cases), but rather is coming from her own will because she also deeply desires to do the opposite. In short, we come back to a parallel processing brain in which separate but interconnected cognitive processes aim at competing goals (different possible choices) represented by competing attractors. The indeterminism that is an obstacle to attaining one of the competing goals (making one of the choices she is faced with) is present because of her simultaneous conflicting desire to attain the other goal. In such circumstances, when either of these cognitive processes attains its goal (which amounts to making a specific choice), it will be like your solving the mathematical problem by overcoming the background indeterministic noise created by the presence of the competing process. 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 could then say this in the present case, whichever one is chosen. For 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 presence of the competing cognitive process.

In short, when thinking about how free will and responsibility might be reconciled with indeterminism, the idea is not to think of the indeterminism involved in free choices as a cause *acting on its own* (in which case we would have “mere” chance), but as an ingredient in a larger goal-directed process or activity in which the indeterminism functions as a hindrance or obstacle to the attainment of the goal. This idea is the role suggested for indeterminism in the cognitive processes leading to self-forming choices. These cognitive processes are temporally extended goal-directed activities in which indeterminism is a hindering or interfering element. The choices that result from these temporally extended activities thus do not pop up out of nowhere, even though undetermined. They are the *achievements* of goal-directed activities of the agent that might have failed, but did not.

Note, in addition, that if indeterminism or chance does play this kind of interfering role in larger processes leading to choice, the indeterminism or chance will not be the *cause* of the choice that is actually made. This follows from a general point about probabilistic causation. A vaccination may hinder or lower the probability that I will get a certain disease, so it is causally relevant to the outcome. But if I get the disease despite it, the vaccination is not the *cause* of my getting the disease, though it was causally relevant, because its role was to *hinder* that effect. The causes of my getting the disease, by contrast, are those causally relevant factors (such as the infecting virus) that significantly *raised* the probability of its occurrence.

Similarly, in the case of the businesswoman’s choice, the causes of the choice she does make (the moral choice or the ambitious choice) are those causally relevant factors that significantly *raised* the probability of making *that* choice from what it would have been if those factors had not been present, such as her reasons and motives for making that choice rather than the other, her conscious awareness of these reasons, and her deliberative efforts to overcome the temptations to make the contrary choice. The presence of indeterminism lowers the probability that the choice will result from these reasons, motives, and efforts from what that probability would have been if there had been no competing motives or efforts and hence no interfering indeterminism. And since those causally relevant features of the agent – which *can* be counted among the causes of the woman’s choice – are *her* reasons or motives, *her* conscious awareness, and *her* deliberative efforts, we can also say that she is the cause of the choice. The indeterminism or chance (like the vaccination) was causally relevant to the outcome, but it was not the cause. This explains why the husband’s excuse was so lame when he said, “Chance broke the table, not me.”

In summary, I have been addressing the third and final objection commonly made by compatibilists (and by Dworkin as well) against traditional incompatibilist conceptions of free will and responsibility. The objection is that incompatibilist free will makes no sense because it reduces either to

---

---

chance or to the mystery of uncaused causes which have no place in the modern scientific picture of the world. I have argued that an incompatibilist free will (one involving indeterminism) need not reduce to “mere chance” or to the mystery of uncaused causes and that it can be reconciled with modern scientific views about human beings. There is much more to say on the subject, but I have tried to indicate in brief terms how this might be done.

In conclusion, I am not what Dworkin calls a “pessimistic incompatibilist” – one who believes free will is incompatible with determinism, but who thinks that incompatibilist free will is impossible, so that no one is ultimately responsible for doing what they do.<sup>21</sup> I agree with Dworkin (and oppose some philosophers who argue to the contrary<sup>22</sup>) that the acceptance of pessimistic incompatibilism would have dire consequences for morality, law, and society. But, if not a pessimist, I am an incompatibilist nonetheless. I think one can be an incompatibilist about free will and agree with many of the good and right things Dworkin says in this book about morality, law, living well, and “creative freedom.” So to end on an ecumenical note: while, it is important for Dworkin to reject the hydraulic model, with its uncaused causes and other mysteries, it is not essential to his overall project in *Justice for Hedgehogs* that compatibilism be true and incompatibilism *in every form* be false.

---

<sup>21</sup> See DWORKIN, *supra* note 1 (manuscript at 142).

<sup>22</sup> See generally, e.g., DERK PEREBOOM, *LIVING WITHOUT FREE WILL* (2001) (arguing that persons can live sensibly and humanely without a belief in free will of the genuine kind that is incompatible with determinism).