Pessimists, Pollyannas, and the New Compatibilism

Paul Russell

The Oxford Handbook of Free Will (1 ed.)
Edited by Robert Kane

Print Publication Date: Mar 2005 Subject: Philosophy, Metaphysics
Online Publication Date: Sep 2009 DOI: 10.1093/oxfordhb/9780195178548.003.0010

Abstract and Keywords

This article examines a number of contemporary compatibilist views on freedom and responsibility. The discussion is organized around themes from Daniel Dennett's influential compatibilist work, Elbow Room (1984), and in the light of these themes the article also considers other compatibilist views, including those of Paul Benson, Martha Klein, John Fischer and Mark Ravizza, Robert Audi, and Kevin Magill. Through these and other authors, the article discusses a variety of topics that have been of concern to contemporary compatibilists, such as control, reflexivity, responsiveness to reasons, “moral luck,” the place of character in moral evaluation, ultimacy, blameworthiness, and normative elements of freedom.

Keywords: freedom, responsibility, Daniel Dennett, control, reflexivity, moral luck, moral evaluation, utlimacy, blameworthiness

> If a man is a pessimist, he is born a pessimist, and emotionally you cannot make him an optimist. And if he is an optimist, you can tell him nothing to make him a pessimist.
>
> Clarence Darrow

THE aim of this chapter is to examine recent contributions to compatibilist literature on freedom and responsibility that are not discussed in the prior chapters of part IV. Although the views of several authors will be considered, discussion will be organized primarily around Daniel Dennett's *Elbow Room*, an important work in the evolution of the “new compatibilism.”

1. Cheerful Compatibilism and the Bogeymen of Pessimism

Dennett's discussion of the free will problem begins with the observation that this is a subject that people care about—it is not simply an intellectual puzzle (p. 230) looking for a solution. One group believes that if determinism is true, and "every deed and decision is the inexorable outcome… of the sum of physical forces acting at the moment," then the human condition would be a "terrible" and "frightening" existence (Dennett 1984: 1–5). Freedom would be an illusion, and we would be reduced to "awful" circumstances similar to those of individuals who find themselves imprisoned or paralyzed, or subject to (hidden) control and manipulation by others.

Incompatibilist views of this kind generate, and reflect, strong emotional responses that can be labeled as "pessimistic." Dennett's fundamental objective in *Elbow Room* is to discredit incompatibilist pessimism and to vindicate a more "optimistic" position (ibid.: 19, 169). According to Dennett, the thesis of determinism has none of these bleak implications for the human condition, and we do not require the metaphysical system building of libertarianism to "ward off non-existent evils" (ibid.: 4; and compare Strawson 1962).[1]

The opening chapter of *Elbow Room* provides a vivid and lively account of how incompatibilist pessimism acquires its psychological grip over us. Our worries and anxieties about determinism, says Dennett, are the product of "fearmongery" by philosophers. It is philosophers who have "conjured up a host of truly frightening bugbears and then subliminally suggested, quite illicitly, that the question of free will is whether any of these bugbears actually exist" (Dennett 1984: 4). The arguments of these pessimistic "gloomleaders," says Dennett, rely on thought experiments that serve as "intuition pumps" designed to produce the same relevant negative emotional response (ibid.: 12, 18). According to Dennett, however, these thought experiments do not so much illuminate the problem as artificially create it by means of misleading analogies and metaphors.

In Dennett's view, the analogies and metaphors concerned "do not in the slightest deserve the respect and influence they typically enjoy" (ibid.: 7). His method in *Elbow Room* is to examine carefully these incompatibilist intuition pumps and to show how they are systematically misleading. In this way, Dennett plays the part of a philosophical therapist, trying to release us from the set of worries and anxieties produced by these misleading analogies.[2] If the therapy succeeds, then the free will problem, as traditionally conceived, "dissolves".[3]

A particularly important subset of the bugbears that Dennett wants to discredit are various "bogeymen," viewed as agents who are really in control of us. The class of bogeymen can itself be subdivided into distinct groups. The first are those analogies that imply that our will somehow fails to govern our conduct, effectively disconnecting us from any (causal) influence on the world. These cases include, for example, imagining ourselves as living in a prison run by an invisible jailer, or being in the clutches of a puppeteer who controls our every movement no matter how we may struggle against him. These versions

of the bogeymen (p. 231) control us not by controlling our will, but by moving our bodies directly and rendering our efforts and preferences inert.

Closely related to these bogeymen are more general worries about fate, the view that all our efforts and deliberations are futile. The concern here is that if determinism is true, and everything that we think and do is governed by causal laws, then we are subject to conditions of universal fate. This bugbear, says Dennett, “looms large” in the free will debate, and the intuition pumps described above do much to support and promote it.

Another subset of bogeymen operate on us in a different way. In these cases the worry is not that our wills fail to guide our conduct, but rather that the way we deliberate and will is controlled by another agent. In these cases, although actions are produced by our will, our will is not truly our own. Examples of this anxiety include cases of hypnosis, or manipulation by an evil neurosurgeon using electronic implants to control us. In such cases we may not even be aware that we are being controlled by another agent. We have the illusion of freedom.

A further worry—in some ways the opposite of the bogeymen anxieties— is that if determinism is true then there is no agent in control at all, since we are really nothing more than mere machines or automata responding in predictable ways to stimuli in our environment. On this view of things, human beings are not much different from simple insects, which can be easily manipulated by more sophisticated beings who control their environment. A wasp, for example, may look as if it makes choices and decisions, but it is really just biological machinery operating according to established causal laws—no real agent is at work. If determinism is true, says the incompatibilist pessimist, then human beings are not much better off than an insect operating in this fashion.

Dennett's objective is to show that all these intuition pumps are, in various ways, misleading. For the purpose of understanding his project, I will focus on his examples of bogeymen and the two different ways that they threaten human freedom. In order to distinguish among the various categories of pessimistic concern, I will introduce a spatial metaphor of distance. Close-range pessimism concerns those cases where the worry is that our will fails to guide our conduct. Middle-distance pessimism is the set of worries we have in circumstances where we believe that we are unable to properly regulate our own will, either because we cannot respond to available reasons or we are subject to manipulation of some kind. I also consider worries that our will is ultimately determined by causal antecedents that we cannot control. I refer to this concern as “pessimism at the horizon.” (See the diagram at the end of this chapter)

(p. 232) 2. Classical Compatibilism and Close-Range Pessimism

A number of Dennett's basic arguments to discredit the bugbears of incompatibilism are taken straight from the shelf of classical compatibilism—as developed by empiricist

thinkers from Hobbes and Hume to Schlick and Ayer. (Classical compatibilism still has distinguished defenders. See, for example, Davidson 1973.)[4] The classical arguments deal primarily with close-range pessimism. The position taken is that the traditional free will debate is a "pseudo-problem," the product of a series of conceptual or terminological confusions. The distinction that is fundamental to this position is that between caused and compelled action. According to this view, free actions are caused by our desires or willings. In contrast to this, unfree actions are brought about by "external" causes, independent of the agent's desires or will. Under these circumstances, the agent is forced or compelled and therefore not responsible for the action. In this way, the classical compatibilist position maintains that free action is to be distinguished from unfree action not by the absence of causes, but rather by the type of causes at work.

Another aspect of the classical position is a diagnosis of the source of incompatibilist confusion on this subject. The "metaphysical" interpretation of the causal relation is supposed to imply that a cause somehow forces or compels its effect to occur. Since freedom is, properly understood, opposed to compulsion, this would imply that an action that is caused must also be compelled, and so unfree. However, when the causal relation is properly understood in terms of a regular succession or constant conjunction of like objects, then all suggestion of causes forcing or compelling effects is removed. To say an action is caused by some antecedent willing by the agent is to say only that events of the first kind regularly follow events of the second kind—nothing more is involved.[5]

The classical compatibilist position also employs the distinctions introduced above to dismiss incompatibilist worries about fatalism. Incompatibilists argue that if determinism is true then all human beings are subject to fate, and any effort to alter or change the future is futile. According to classical compatibilism, this simply confuses two distinct issues.[6] Determinism is the thesis that everything that occurs, including our deliberations and decisions, are causally necessitated by antecedent conditions. Fatalism, by contrast, is the thesis that our deliberations and decisions are causally ineffective and make no difference to the course of events. Although there may be particular circumstances when we find that our efforts are futile ("local fatalism"), nothing about the thesis of determinism implies that this is the universal condition. On the contrary, as Dennett puts it, "deliberation is (in general) effective in a deterministic but nonfatalistic world" (Dennett 1984: 106).

(p. 233) Moral freedom, as the classical compatibilist understands it, involves being able to act according to the determination of our own will—that is, doing as we want to do or as we please (Hobbes 1962:1, 66–8; Hume 1955: 95). On this account, therefore, freedom is a matter of freedom of action, the absence of any external impediments or obstacles. Accompanying this positive doctrine is the negative thesis that incompatibilist attempts to provide some account of free will, as distinct from free action, are radically mistaken and confused. More specifically, the notion of free will, it is claimed, is simply meaningless and absurd (Hobbes 1962: 1, 61–62). The only freedom that we need or want, according to this view, is to be able to guide our conduct by means of our own desires and willings. Any effort to go beyond this and explain moral freedom in terms of control over our own

will inevitably leads to either metaphysical obscurity or the absurdity of an infinite regress.

3. Reason, Self-Control, and Middle-Distance Pessimism

On the face of it, the classical compatibilist arguments deal effectively with close-range pessimist worries about being unable to regulate conduct through our own will. A determined world should not be assimilated to conditions of an invisible jail or being a puppet, since we can still distinguish circumstances where we act according to our will from those in which we do not.

These observations and reflections, however, fall far short of dealing with middle-distance pessimism. The most obvious difficulty facing any conception of moral freedom identified with the ability to act according to the determination of an agent's desires or willings is that such freedom is something that an animal, a child, or a mentally ill person might enjoy—all paradigmatic cases of individuals who lack moral freedom. Related to this point, some individuals, such as the kleptomaniac, appear to act according to compulsive desires. In cases of this kind, the agent's desires constitute *internal* obstacles to doing what the agent (reflectively) truly wants to do. Clearly, then, classical accounts of freedom understood simply as free action cannot draw the sorts of distinctions that we need to make in this sphere.[7]

These familiar incompatibilist objections to classical compatibilist accounts of freedom seem closely related to some of the worries raised by Dennett's "bogeymen." In the case of middle-distance pessimism, the concern is not that our will does not guide our behavior, but rather that we are unable to regulate our will (p. 234) according to reason or our own true values. Two of Dennett's examples speak directly to this problem—hypnotism and manipulation through neurological implants. The specific way that we interpret these cases, and the worries associated with them, will shape the way we judge the prospects of the "new compatibilism."

Dennett's interpretation of these cases, and the fears that they generate, center on two closely related issues. The first concerns the worry that we are not able to regulate our will in light of reasons that are available to us. The second is that our will is in some way being manipulated by another agent, and so our conduct is being indirectly controlled through control of our will. Under these circumstances our conduct reflects, not our own reasons and interests, but rather those of our manipulator. If Dennett can show that determinism has none of these unpleasant and disturbing implications then, he believes, he has discredited middle-distance pessimism.

The first step in his approach is to explain the nature of the relationship between our capacity for reason and the kind of freedom that is worth wanting. What we want, says Dennett, is to be the sort of creatures that are able to be "moved by reasons" (Dennett 1984:

25). Our reasons for acting are interpreted in terms of our fundamental interest in "self-preservation" and "self-replication." As finite beings, of course, our ability to represent all such reasons to ourselves is limited, but this does not mean that our sensitivity to relevant changes and variations in our environment is not significantly greater than that of other creatures. What is especially important to us, Dennett argues, is our ability to consider not only the direct objects of our desires, but also to reflect on our beliefs and desires themselves. This kind of reflective capacity enables us to question the evidential credentials of our beliefs, as well as the soundness and coherence of our desires. This constitutes, Dennett suggests, "a major advance in the cognitive arms race" (ibid.: 37). (This account of our reflective capacity is, of course, closely related to other "hierarchical" or "real self" theories of freedom, as advanced by, for instance, Frankfurt 1971 and Watson 1975.)

According to Dennett, the particular importance of this "power of reflexive monitoring" is that it helps us to deal with worries about manipulation by others. An agent who is able to examine and monitor his own beliefs arid desires will detect "abnormalities" in their causation (ibid.: 1984: 30). With this ability, an agent can unmask "sneaky manipulators" or "evil tricksters"—which makes it difficult to trap him in disturbing situations of the kind suggested by middle-distance bogeymen. These abilities to self-monitor and escape the clutches of (evil) manipulators evolve and develop naturally and gradually—both in the individual and in the species. Nothing about the thesis of determinism suggests that we do not possess and exercise such abilities. What is crucial, however, is that we do not allow ourselves to be deceived by "intuition pumps" that conceal the complexity of our rational and reflexive powers. For the purposes of understanding human freedom, Dennett argues, *complexity matters* (ibid.: 12, 34, 37–38).

(p. 235) Central to clarifying the nature of freedom—and escaping our worries about bugbears—is recognizing that what we want or value is control. "We want to be *in control*" says Dennett, "and to control both ourselves and our destinies" (ibid.: 51, Dennett's emphasis). Any individual who is a "controller" must have states that include desires about the states of the "controllee," which must in turn have a variety of states that it can be in (ibid.: 52). Dennett uses the example of controlling an airplane to illustrate this point. By means of anticipating or predicting future states of the airplane, we can keep control of it. There are limits to the range of things that we can do with the plane (that is, degrees of freedom with respect to it). Nevertheless, if we judge things correctly, we can retain control over it. When it comes to self-control, this is what distinguishes us from "mere puppets." We are not helpless in using our foreknowledge and powers of deliberation to "take steps to prevent, avoid, preempt, avert, harness or exploit" wanted or unwanted circumstances. This power of control and self-control is what we want and value. Like the pilot of a airplane, we want to leave ourselves a "margin for error"—lots of "elbow room"—so we can keep control of the situation and do the things that we want to do (ibid.: 62–63).[8] Self-knowledge is essential to maintain and expand this freedom. While not "absolute" or unconditional, human beings enjoy a considerable amount of this kind of con-

trol. One implication of this understanding of control is that there are “degrees of freedom” (ibid.: 53; and compare Bernard Williams 1986: 5).

This account of freedom, as explained by Dennett, clearly goes well beyond the simple definitions suggested by classical compatibilism. On this account, it is not meaningless or absurd to say how free agents are able to control and regulate their own desires and wills. Our powers of reflection enable us to monitor our beliefs and desires, and, when necessary, to detect and “disconnect” unwelcome manipulators. Accompanying this positive doctrine, there are important negative theses about the nature of human freedom. First, a freedom that implies an ability to make arbitrary or causeless decisions or choices is not worth wanting, and not what we actually care about (Dennett 1984: 2). Second, and relatedly, the kind of freedom that Dennett has described does not presuppose that agents “could have done otherwise.” This claim is particularly controversial, although it is consistent with Harry Frankfurt's well-known critique of the doctrine of “alternative possibilities” (Frankfurt 1969).

Dennett endorses Frankfurt's strategy but also argues that it is “insufficiently ambitious” (Dennett 1984: 132). In the first place, Dennett argues, a person may truly state that he could not do otherwise, but not in order to disown responsibility (ibid.: 133–35). Beyond this, if such a condition had to be satisfied to establish responsibility—that is, the agent could have done otherwise under the exact same circumstances—we could never know whether the agent was really responsible, given the epistemological difficulties involved. Finally, not only is the “traditional metaphysical question unanswerable”; even if we knew the answer, it would be (p. 236) useless. We want to know whether the agent is likely to repeat similar kinds of (undesirable) conduct again—and to know this we do not need to know if she actually had “alternative possibilities” available to her under the specific circumstances of her action. The question that matters to us is whether or not a flawed character trait needs to be corrected (ibid.: 137–38).

There is, according to Dennett, another insidious (middle-distance) bugbear that needs to be exorcised from the overactive incompatibilist imagination—one with, he believes, an especially powerful hold over us: the worry that if determinism is true then we are (somehow) “controlled by nature” or “controlled by the past” (ibid.: 50, 61, 72). This way of presenting the pessimist's anxieties does not rely on any fictional or hypothetical case of (evil) hypnotists or neurosurgeons at work. On the contrary, the source of the anxiety seems much closer to traditional theological worries about God's omnipotence and omniscience undermining the possibility of human freedom. Clearly God is not conceived of as evil, but vis-à-vis our aspiration to be true self-controllers, God may be viewed as a kind of cosmic bogeyman. In the secularized/naturalized version, however, the role of God is played by “Nature” or the “Past,” but the same general worry persists: while we appear to be self-controllers, control nevertheless slips away through the causal chains to an external and alien source. Self-control, therefore, is really an illusion.

Dennett's reply is that such worries rest on simple confusion about the nature of control. To be a *controller*, as we have noted, involves being an agent with desires that can drive

the controllee into some preferred state or another. The controller must also receive "feedback signals" from the object if it is effectively to control it (ibid.: 72). All talk of being controlled by Nature or by the Past plainly involves *personification* (ibid.: 57, 72). Without this, these bugbears disappear—neither Nature nor the Past can properly be said to be "controllers" of any kind, whether determinism is true or not. On Dennett's account this (basic) confusion about the nature of control motivates much of the incompatibilist's pessimism and accompanying resistance to the thesis of determinism.[9]

This analysis of incompatibilist worries covers three issues that we should carefully separate: (i) Do human purposes and choices have determining causes that ultimately originate externally (2) Is the ultimate source of our purposes and choices another intentional agent, who is in control of us? and (3) If there is such an agent in control of us, is the quality of its moral character good or evil? Pessimist anxieties, according to Dennett, depend largely on the last two issues. It is especially horrible to imagine ourselves under the control of another demonic or evil agent (for example, "hideous hypnotist" and the like). Nevertheless, even a benevolent controller, looking out for our interests, leaves us with a sense of chill, since there remains the fear that some other agent is "really in control of us."[10] When we consider the first issue by itself, Dennett maintains, we have no reasonable basis for being troubled or disturbed by the thought that the ultimate origins of our deliberations and choices lie outside of us.

(p. 237) Dennett associates worries about the ultimate origin of our deliberations and choices with the aspiration to "absolute agenthood"—to be a perfect, Godlike self-creator (ibid.: 83–85). It is his position that this aspiration is both impossible and unnecessary, since it is not needed for the kind of freedom that we care about (that is, "self-control" as he interprets it). The incompatibilist view is that, contrary to Dennett, worries about ultimacy or "absolutism" are essential to our conception of ourselves as true self-controllers, and libertarians maintain that this kind of freedom (which rules out determinism) is something that human beings are actually capable of. The distinct set of worries associated with ultimacy are the basis of "pessimism at the horizon." The critical question that faces us is whether Dennett is justified in dismissing these concerns at the horizon as both incoherent and unnecessary.

4. Middle-Distance Refinements and Difficulties

It is clear that Dennett's version of the new compatibilism involves a number of controversial claims. At this stage, however, I want to consider some interesting amendments and modifications that have been suggested in two essays by Paul Benson. In "Freedom and Value" Benson argues that free agency requires another "equally significant ability" apart from control, the ability "to appreciate values." More specifically, to attribute free agency correctly in a given context depends, according to Benson, "partly on the content of the agent's normative understanding, not just on the agent's having some valuational point of view or other" (Benson 1987: 472). Benson maintains "that obstacles to compe-

tent appreciation of the norms that apply to our actions are as much impediments to full freedom as are certain obstacles to the expression of our evaluative judgments in our will or certain obstacles to the realization of our will in our conduct" (ibid.).

Benson points out that the omission that he is concerned with in compatibilist accounts of moral capacity (that is, normative competence) is addressed in Susan Wolf's essay "Asymmetrical Freedom" (Wolf 1980; and see also idem 1990), but he argues that what is missing from her account "is any discussion of why specifically *freedom* involves the competent appreciation of value" (Benson 1987: 474). To answer this question, we need to reflect on why the power of control is so important to us. We care about control, Benson suggests, because we care about the values by which our actions are assessed. This, in turn, reflects our "deep-seated desire to be able to justify our conduct" (Benson 1987: 475; and compare (p. 238) Scanlon 1988: 170–72). Since the norms governing our actions are important to us, so too must be the ability to regulate our conduct by means of our evaluative judgments. Benson continues:

> [*I*]f we care deeply about the value of our actions, then we want more than the power to translate our own value judgments into effectual willing. We also want to be able to appreciate the relevant values and arrive at competent appraisals of the alternative courses of action we face. Our concern for those values would be practically impotent if we could not bring them competently to bear in our deliberations about what to do. (Benson 1987: 475)

Benson uses these observations about the importance of normative competence to shed light on another feature of fully free action that is intimately connected with it: the "enduring belief that a completely free act is fully our own" (ibid.).

Free acts are fully our own, Benson argues, "only insofar as they potentially afford appropriate bases for normative assessments of us in face of which we have no excuse" (ibid.: 482). When we lack any control over what we do (for example, cases of compulsion), the action provides no basis for "moral disclosure" and thus cannot be fully our own. Similarly, when agents lack normative competence, Benson argues, their conduct cannot reveal their moral values and so cannot disclose what they are like as persons in the relevant respect. The incapacity involved may be severe enough to render the individual wholly ignorant of normative standards and when and how they apply (as in the case of infants or severe mental illness). In other cases, the agent may adequately *appreciate* the pertinent values but cannot use their normative insights to *regulate* or guide their conduct (for example, older children, the severely deprived, and so on). (For a different cornpatibilist perspective on the issue of deprivation and blameworthiness, see Klein 1990: esp. ch. 4, sec. 3. For another view closer to Benson's, see Wallace 1994: 231–35.) The general point, in all such cases, is that actions coming from agents who lack normative competence cannot reveal their moral values and, as such, cannot be said to be "fully their own."[11]

In a more recent essay, "Free Agency and Self-Worth," Benson modifies his position. He argues, in this context, that the "normative-competence condition" is too strong, insofar as it is "content specific." That is, Benson now accepts the view that "any desires, plans, values, beliefs, etc., can be involved in the motivation of free action"—free agents must be

able to “commit themselves to whatever motives they please” (Benson 1994: 653, 663). On the new account, Benson refuses “to restrict substantively persons’ desires, values, life plans or normative capacities in the name of freedom” (ibid.: 665; compare Christman 1991b): 356–59).[12] However, this more “permissive” position is not wholly “neutral” about content. The weaker condition that Benson now advances is a “self-worth condition.” Free agents must “have a certain sense of their worthiness to act, or of their status as [competent] agents, which is not guaranteed by their abilities to act freely” (Benson 1994: 650).

(p. 239) The condition of self-worth, Benson argues, helps us to understand a variety of cases where agents do not face any of the “standard impediments” to free agency but are nevertheless not fully free. Among the cases that he cites are the effects of severe shaming and slavery, conditions that undermine a person's confidence in their own competence as an agent and, as such, constitute an assault on their sense of “moral dignity as persons.” One particularly important aspect of this condition is that it draws our attention to the “social dimension of free agency” (ibid.: 661). Related to this point, this condition of free agency also clarifies that the value of free agency lies in part with “our sense of being in a position to answer for [our] conduct,” which is itself “partly constitutive of [our] sociality” (ibid.: 668). “A blow to our freedom,” Benson argues, “can obstruct our ability to express through our conduct who we are, but it can also be a blow to our sense of who we are as social creatures” (ibid.: 668).[13]

Another important set of issues that arise from Dennett's discussion concern the question of how freedom relates, in more precise terms, to our capacity to be guided by reason. Recent work by John M. Fischer and Mark Ravizza (1998) provides an influential and illuminating discussion of this problem. Fischer and Ravizza make clear that the relationship between “reason-responsiveness,” on one side, and freedom and responsibility on the other, is open to very different interpretations. On the account that they provide, our capacity to respond to reasons depends on our (natural) “human deliberative mechanisms” (ibid.: esp. 34-41). A free agent, on a “strong” interpretation, operates with a mechanism that is *always* receptive and responsive to available reasons. Under these circumstances, the agent's reasons, choices, and actions reliably “track value” or “the reasons there are” in every case (Fischer and Ravizza 1998: 42; compare Nozick 1981: 317–62). Clearly, however, this condition is too demanding, since we would then be unable to hold an agent responsible when “tracking” reason *fails*. So what is required is a weaker theory that can accommodate cases where the (actual) mechanism fails, as well as cases where it succeeds.

Fischer and Ravizza employ considerable ingenuity trying to develop a “weaker” or “moderate” account that can deal with worries of this kind. A plausible account, which can serve the purposes of compatibilism, must describe “mechanisms” that can fail under some conditions, without being *systematically* unreliable (that is, too “weak”). We need, therefore, some principled way of distinguishing and identifying mechanisms that are sufficiently reliable in responding appropriately to reasons. When it comes to recognizing what reasons there are (that is, receptivity), there must be, Fischer and Ravizza argue, some appropriate *pattern* of reason-receptivity. That is to say, “the actual mechanism that

issues in [the agent's] action must be at least '*regularly*' receptive to reasons" (Fischer and Ravizza 1998: 70–1). This avoids the worry that the mechanism in question could be reason-receptive in an isolated case but otherwise fails systematically. When it comes to choosing in accordance with the available reasons (that is, reactivity), (p. 240) however, Fischer and Ravizza argue that the (stronger) demand for regular-reactivity or a pattern is not required. All that needs to be satisfied:, they maintain, is the weak condition that in a given case the mechanism has been shown to be reactive to available reasons (Fischer and Ravizza 1998: 73–76).[14]

This account of "moderate reason-responsiveness" introduces an "asymmetry" between the receptivity and reactivity requirements. Fischer and Ravizza describe this as follows:

> In the case of receptivity to reasons, the agent … must exhibit an understandable pattern of reasons-recognition, in order to render it plausible that his mechanism has the "cognitive power" to recognize the actual incentive to do otherwise. In the case of reactivity to reasons, the agent … must simply display *some* reactivity, in order to render it plausible that his mechanism has the "executive power" to react to the actual incentive to do otherwise. In both cases the pertinent power is a general capacity of the agent's mechanism, rather than a particular ability of the agent (i.e., the agent's possession of alternative possibilities—the freedom to choose and do otherwise). (Fischer and Ravizza 1998: 75, emphasis in original)

Two (related) difficulties arise from these claims. The first problem is that it is unclear what justifies the "asymmetry." If a "pattern" or "regularity" is needed for receptivity, why is this not the case with reactivity? Clearly, Fischer and Ravizza hold that strengthening the reactivity requirement, in line with the receptivity requirement, would be too demanding, since we do not want to excuse agents whose mechanism is regularly receptive and has shown that it can react to reason. The controversial assumption that this position rests upon is that "reactivity is all of a piece in the sense that the mechanism can react to all incentives, if it can react to one" (Fischer and Ravizza 1998: 73–74). It may be argued, however, that this same reasoning can be applied to the receptivity requirement, which would result in a return to a "weak reason-responsive" view. On the face of it, therefore, the asymmetry that Fischer and Ravizza introduce, in order to arrive at a "moderate" position, seems to depend on ad hoc adjustments rather than principle-driven considerations.

There is, I believe, an even more fundamental difficulty for a reason-responsive view of the kind that Fischer and Ravizza seek to defend, The objection may be raised that it is unclear how the mere possession of such reason-responsive mechanisms or capacities can render agents sufficiently in control, unless they also have control over how the capacity is actually *exercised* within the particular conditions of action. On this view of things, the responsible agent needs more than simply the general capacity for reason-responsiveness (under some interpretation). What is also needed is a capacity of *exercise control*, which means that the agent is able to direct the specific way that her capacity for rational self-control moves her.

(p. 241) Any attempt to satisfy this demand is, of course, liable to lead us back into the conundrums associated with "ultimacy" and "absolute agency" (as discussed later in this

chapter). While it may well be that exercise control is a demand that can never be satisfied, it will not suffice for the compatibilist to argue this point— since the "moral skeptic" or "hard determinist" may agree about this. The point that the compatibilist needs to establish is that exercise control is *unnecessary* for responsibility, and that the mere possession of powers of rational self-control will suffice. (For an interesting, although I think unsuccessful, attempt to make this case, see Wallace 1994: 180–93; and 161–62, 201–14.)

The difficulties that we have been considering relate primarily to the possibility that reason-responsive mechanisms may sometimes fail to respond appropriately to available reasons, without excusing the agent. There are, nevertheless, also difficulties associated with "strong" mechanisms that cannot fail (that is, always "track value"). In cases of this kind, since the agent is guided flawlessly by reason and enjoys perfect practical reason, she may be viewed as perfectly free. This view, however, does not entirely square with all our intuitions on this subject. More specifically, it may be argued that an agent who is *naturally* governed by (moral) reason, and so does what is required of her *effortlessly*, does not *deserve* moral praise. Moral praise should be reserved for those who must "struggle" to be good and do the right thing. Certainly, this claim captures the spirit of important strands of neo-Kantian incompatibilism (Campbell 1951: 130–33). However, some compatibilists, such as Martha Klein, embrace this view and have made it an essential element of a compatibilist approach to moral responsibility (Klein 1990: 167–71; compare Wolf 1990: 138–42).[15]

The general point that these observations bring to light is that reflection on both the success and failure of reason-responsive mechanisms present compatibilism with difficulties, and the relationship between rationality and freedom is by no means straightforward. Dennett's tendency to present incompatibilist concerns as based on confusion and exaggerated worries of various kinds leads him to underestimate the (genuine) difficulties and obscurities involved in articulating a plausible compatibilism as it relates to middle-distance issues.[16]

Nevertheless, while significant "gaps" in Dennett's compatibilist position are apparent, it is evident that he succeeds in outlining how compatibilists can deal with middle-distance worries about self-control, as they relate to questions of rationality and manipulation. Moreover, as Dennett's analysis of the "problem cases" suggests, these two categories are intimately connected, since cases of manipulation can be understood as "problematic" precisely because they involve a break-down in the agent's sensitivity to reasons. (See also Wallace 1994: 175–77, for a related account of how such "problem cases" can be interpreted in terms of a breakdown of rational self-control.) Dennett's strategy is to argue that our (natural) *complexity*, not indeterminism, provides us with the ability to be sensitive (p. 242) to available reasons and to guide our conduct on this basis. The same general ability gives us powers of "self-monitoring" that enable us to detect and escape from (threatening) manipulators. These incompatibilist bogeymen., therefore, need not frighten us anymore.

5. Ultimacy and Pessimism at the Horizon

Middle-distance pessimism, as we have seen, is generated by worries associated with intuition pumps and bogeymen that imply that we are somehow unable to regulate our will according to reason and what we reflectively care about. This is why we find (hypothetical) cases of manipulation disturbing: we want our will to respond to reason and we do not want another agent to control our will (in service of alien interests or reasons). Dennett maintains that in order to avoid these worries we do not need to be "absolute agents" capable of self-creation ex nihilo. More specifically, it is a false dilemma to suggest that either we are "a completely self-made self, one hundred per cent responsible for its own character" or we are "mere dominoes" in the causal chain (Dennett 1984: 100, 156–57). All that we want, says Dennett, is "to be as immune as possible from manipulation and dirty tricks and as sensitive as possible to harbingers of future vicissitudes that might cause us to alter course in the right ways—so that we can face the world with as much elbow room (as large a margin for error and as little relevant uncertainty) as we can get" (ibid.: 72–73).

Dennett refers to a number of philosophers who have presented objections that are supposed to show that our worries about determinism extend to issues on the horizon (ibid.: 33, 75, 83–84). He cites, for example, A. J. Ayer's description of "implanted" desires and beliefs (Ayer 1954: 9); Paul Edwards's observation that if determinism is true then even our efforts at self-creation must be "the result of factors that are not of [our] making" (Edwards 1961: 121); and Thomas Nagel's worries about "luck" as it concerns even "the stripped-down acts of the will itself" (Nagel 1979:183). Each of these critics raises variations on the problem of ultimacy. For the purpose of this essay, however, I turn to Martha Klein's particular account of this problem.

Although Klein defends a ("partial") compatibilist position, she maintains, nevertheless, that our ordinary moral intuitions support certain incompatibilist claims (Klein 1990: 3 and ch. 4).[17] More specifically, according to Klein we generally accept "that one of the things which disqualifies an agent from blameworthiness (p. 243) is his not having been responsible for the causes of his decisions or choices" (ibid.: 51). This conviction commits us, she says, to a "U-condition" for agent accountability: the condition that "agents should be ultimately responsible for their morally relevant decisions or choices—'ultimately' in the sense that nothing for which they are not responsible should be the source of their decisions or choices" (ibid.: 51).[18] Klein's interpretation of the basic rationale behind the U-condition is that if agents acts are caused by factors for which they are not responsible, it is not obvious how they can be responsible for acting as a result of those factors (ibid.: 50). (This way of interpreting the U-condition and its significance is open to revision. See, in particular, Kane 1996a: esp. chs. 3 and 5; and also the essays by Kane and Galen Strawson in this volume, chs. 18 and 19.)

In support of the U-condition, Klein cites a number of "problem cases" that closely resemble Dennett's "bogeymen" examples (Klein 1990: 70–75, 89–90). These include victims of brain tumors, implantations, brainwashing, and hypnosis. The feature these cases share,

Klein maintains, is that in each the agent's decisions can be traced to causes for which he is not responsible, and so he ought not to be blamed (ibid.:70). The example of the brain tumor is especially important to Klein's case for the U-condition, because it highlights the point that the real source of concern is not the "implantation" of desires and beliefs by others, but rather that the agent is not the true source or origin of his own motivations, since "he did not choose (to have) these states of mind" (ibid.: 73).[19]

Klein extends this reasoning and applies it "to the relatively pedestrian and non-threatening-sounding causes of genetic endowment and environment." The U-condition theorist reasons, says Klein, that since the agent "is no more responsible for his genetic endowment and upbringing than he is for the designs of a malevolent demon or brainwasher," it follows that he "is no more responsible for a personality which (perhaps) depends on his brain in a normal state, than he is for the personality change attributable to the brain tumour" (ibid.: 75). From the perspective of the U-condition advocate, unless this condition is met, it will simply be a "matter of luck" whether or not an agent's will is governed by "good" or "bad" desires (ibid.: 165-66). Under these circumstances it would be *unfair* to impose unpleasant treatment such as blame and punishment on an agent who is the (undeserving) "victim" of bad desires.

Dennett's initial line of reply to these worries is that his observations on middle-distance pessimism, and the bogeymen that it conjures up, discredit Klein's concerns about "ultimacy" or "absolute agency." Take, for example, worries that we may have about "implantation" of desires and beliefs. According to Dennett, so long as the agent possesses the relevant degree of "complexity" to be capable of self-monitoring, then she will be able (eventually) to unmask "the process of conditioning" (Dennett 1984: 33–34). Of course, if this capacity is destroyed or damaged by the conditioning process, then the agent is not a self-controller in the full sense of the term—but determinism itself does not imply this. What is (p. 244) worrying about brain tumors is not fears of manipulation by others, nor that our thoughts and actions are caused, but rather that we may become insensitive to reasons and consequently act in irrational and unpredictable ways (compare ibid.: 64-65). While *this* is frightening, there is no basis for supposing that determinism implies it. In sum, we do not need "absolute agency," says Dennett, to avoid the sorts of worries that Klein's "problem cases" present to us.

According to the U-condition theorist, this general line of reply entirely misses the point. It is not denied that agents may possess some relevant capacity to be "reason-responsive" and to revise and alter their character on the basis of reflection. We might well be able to distinguish agents of this kind from individuals who lack these capacities (as new compatibilism suggests). Nevertheless, all this only postpones the fundamental difficulty. While our beliefs and desires may be subject to self-monitoring activities of various kinds, it remains true that these activities are themselves conditioned by factors that are not of the agent's own making.[20] Reflection on this process, therefore, strips away our confidence that we are truly "self-creators" *even in the normal case*. For this belief to be sustained, we must presuppose some power to undertake "self-forming actions" that enable us to be the (ultimate) origin of our character and conduct.[21] The sorts of capacities that Dennett

and other new compatibilists in his mould describe fall short of this, and so their strategy fails to relieve pessimistic worries at the horizon.

Other lines of reply, however, are still available to Dennett. The first is to argue that many of these worries are motivated by confusion about "luck." It is simply a mistake, he claims, to suggest that individuals who are self-controllers of the kind that he has described are subject to "luck" because they fail the test of "absolute agenthood." These individuals are not "just lucky," he argues; they are "skilled" and "gifted" members of "the community of reason-givers and con-siderers" (ibid.: 92–100). When we identify individuals with these abilities we do not—and should not—treat them as simply "lucky" or "unlucky." On the contrary, we provide them with reasons and treat them accordingly.

This response, I believe, fails to confront the real worries that the U-condition presents. Without ultimacy, two crucial modes of control are absent: (1) The actual "reason-responsive mechanisms" that we possess are *acquired* in ways over which we have no final control (in both the normal and abnormal case). The character of these mechanisms, however, plainly determines the kind of choices and decisions that we will actually make.[22] (2) Apart from worries about how we acquire our (given) reason-responsive capacities, we may also worry about our ability to control the way that these capacities are *exercised* in specific circumstances (as discussed earlier). If determinism prevails, then the way capacities for self-creation and self-monitoring are exercised in a given situation will ultimately be determined by factors the agent cannot control.[23] Dennett is clearly right to assert that (p. 245) this does not reduce us to the condition of a "domino" or "zombie" and so on, but it is still true that without ultimate or absolute agency of some kind we lack these vital modes of (self-) control. It may be argued, therefore, that Dennett is too complacent in face of these problems, and consequently his "considerable optimism" (ibid.: 48) has the same pollyannish appearance that plagues classical compatibilism.

Dennett has, nevertheless, more cards to play. Up to this point his methodology has been faithful to the aims of "descriptive metaphysics."[24] That is to say, his position has been that our everyday attitudes and practices associated with moral freedom and responsibility are not threatened by any (confused) pessimist worries at the horizon. This is consistent with Dennett's "ordinary language" effort to expose the "bugbears" and "bogeymen" for what they really are—artificial creations of professional philosophers in the Western tradition. However, when it comes to dealing directly with worries at the horizon as they relate to issues of responsibility, Dennett takes a sharp turn in the direction of "revisionary metaphysics".[25] The argument here is that worries about ultimacy may be motivated by a conception of *responsibility* that, although deeply rooted in the Western philosophical and theological tradition, is nevertheless hopelessly incoherent and implausible—and so ought to be jettisoned. What really sustains "absolutism," on this view, is an understanding of responsibility that is committed to a conception of "total, before-the-eyes-of-God Guilt" (ibid.: 165-66; on related themes see also Bernard Williams 1986). An absolutist conception of *desert* of this kind takes issues of responsibility out of the relevant

(human) practical contexts that should concern us and tries to place them on metaphysical foundations that are disconnected with these legitimate and intelligible concerns.

In opposition to the absolutist view, Dennett prefers a conception of responsibility that is thoroughly utilitarian and forward-looking, and he leans heavily on “engineering” metaphors when describing how this system operates (Dennett 1984: chs. 6, 7). Responsibility, he argues, should be understood in terms of “the rationale of punishment,” and its rationale is to support the criminal laws of society. That is to say, we punish individuals when we think they are “mentally competent” enough to be deterred or reformed by the threat or imposition of sanctions. All this is not only a highly “revisionary” approach; it also takes a (large) step back in the direction of classical compatibilism.

Although the utilitarian features of Dennett's position are very familiar, a more unusual and interesting aspect of his discussion draws attention to the question of how responsibility and character are related—a subject that is generally treated lightly in free will literature. The view that Dennett defends is that in the realm of responsibility, what really interests us is what an action reveals about the character of the agent. More specifically, what we want to know is what we can *expect* from the agent in the *future* (ibid.: 137–38; compare Smart 1961: 300–305). (p. 246) Isolated actions may be “regrettable,” but they are only of moral interest to us insofar as they suggest ways that we can “redesign” agents so they will avoid future “errors” (Dennett 1984: 139–44). The importance of action, on this view, is that it allows us to identify character flaws that can be corrected by means of some relevant sanctions. Actions that do not serve this purpose can be dismissed as “don’t cares”—that is, as cases that it is “rational to ignore” (ibid.: 141).

This view is plainly at odds with our ordinary moral assumption that agents are no less responsible for out-of-character action than for action that is in character. This certainly suggests that Dennett's “revisionism” is more radical than he acknowledges. Beyond this, the critic may also argue that, given that out-of-character action is still produced by the agent's own will, it is entirely reasonable to attribute such conduct to this agent, even if he is unlikely to repeat it in the future (compare Foot 1957: 105–6). Action that is produced through the agent's own will should not be treated the same way as action produced by another agent, or no (moral) agent at all.

The compatibilist can, of course, agree with Dennett that we ought to take the issue of responsibility for character more seriously, without endorsing his forward-looking, utilitarian perspective. Robert Audi has argued, for example, that agents can be held responsible for their character traits, but that this depends on the fact that their character is in some way generated or retained by more basic acts. According to Audi, “all (normative) responsibility traces to acts and ultimately to basic acts,” (Audi 1991b: 307) because a person cannot be responsible “for something over which one has no control” (ibid.: 312). We can be responsible for our traits of character, therefore, only because we have control over our actions, which in turn affects our acquisition or retention of traits (ibid.: 312–3). A view of this kind, Audi maintains, can account for responsibility for character, consistent

with compatibilist commitments, but without utilitarian commitments of the kind that Dennett embraces (ibid.: 319).[26]

Although Dennett gives considerable attention to the question of control and "self-creation," he is not committed to Audi's view, that responsibility for character requires the agent to have control (either generative or retentive) over it. On the contrary, an agent's character could be "implanted" or "conditioned" in ways she could not control, and yet it may still be true that sanctions or moral engineering will be effective in altering or changing her future conduct in desirable ways. Clearly, then, Dennett's pragmatic, utilitarian approach severs any (assumed) link between control and responsibility for character.[27]

My analysis reveals a deep tension in Dennett's entire project in *Elbow Room*. On his account, the relevant authorities or powers in society can (and should) use the conditioning influence of rewards and sanctions to control the character of others. In this way, even though the individuals concerned may possess rational and reflective capacities, in a (deeper) sense they may be truly described as "selves-made-by-others." (p. 247) The irony in all this is that Dennett's pragmatic, engineering approach to responsibility allows real worries about manipulation and "conditioning" to resurface. (There is, indeed, something of the spirit of B. F. Skinner's *Walden Two* to be found in his views on this subject.) To this extent, the first part of Dennett's project, which aims to relieve us of pessimistic anxieties about manipulative "bogeymen," is undermined by the second, which defends a conception of responsibility that places heavy emphasis on the benefits of "social engineering." (A good discussion of why we should be troubled by circumstances of this kind is presented in Kane 1996a: 65–70, 201–4).

In a review of *Elbow Room*, Gary Watson suggests that Dennett's "treatment of responsibility is the least instructive part of the book," and that the weaknesses of his general position are well illustrated by P. F. Strawson in his important essay "Freedom and Resentment" (Watson 1986: 522; and compare Dworkin 1986: 424). A central theme of Strawson's essay is that compatibilists or "optimists" who emphasize only forward-looking, utilitarian considerations in their account of moral responsibility leave an important "gap" in their position. More specifically, according to Strawson, conditions of responsibility must be understood in terms of our natural disposition toward "reactive attitudes and feelings" or "moral sentiments." Such responses to the good or ill will that we detect in the conduct of our fellow human beings are an "essential part of moral life as we know it" (P. F. Strawson 1962: 23). To a limited extent, we can suppress these reactions in particular cases or circumstances: there is no possibility however, that we can *systematically* abandon or suspend our commitment to the whole "complicated web of attitudes and feelings."

These observations, Strawson argues, are highly significant for the free will debate because they reveal what is wrong with both (classical) compatibilist optimism, as well as incompatibilist pessimism. Pessimists are right in saying that a purely utilitarian approach to responsibility leaves out "something vital in our conception of these practices" (ibid. 23). It is a mistake, however, to conclude on this basis that what is re-

quired to fill this “lacuna” in the optimist account is some form of libertarian metaphysics that involves denying determinism (ibid.: 23–25). Contrary to the pessimist, Strawson argues, no theoretical belief in the truth of determinism could lead us to abandon our commitment to the moral sentiments (ibid. 18: compare 10, 12). To suppose otherwise is “to over-intellectualize the facts” (ibid.: 23). When the role of moral sentiment is allowed its proper place in moral life, we can avoid both a crude utilitarian account of responsibility that is divorced from psychological reality, while at the same time avoiding the “panicky metaphysics” of libertarianism. Our sense of *desert* is founded, not on (incoherent) beliefs about undetermined conduct, but rather on the natural, emotional responses that are essential to human life as we know it.

(p. 248) A number of Strawson's followers have picked up on his “naturalistic” arguments and developed his twofold critique of utilitarian optimism, on one side, and of pessimistic worries at the horizon on the other side. (See Ishtiyaque Haji's essay, ch. 9 of this volume, for further discussion of Strawsonian strategies.) Among these contributions to Strawsonian themes is the work of Kevin Magill, who advances arguments that are relevant to Dennett's “revisionary” views about responsibility. Magill maintains, in line with Strawson, that we must resist the temptation to provide a general “justification for punishment, desert and moral responsibility.” The “impulse” to do this, he claims, is based on the (misguided) assumption that a utilitarian principle can be applied to a sphere where a distinct and independent retributive principle operates (that is, that the guilty should suffer). According to Magill, both the utilitarian and retributive principles are “foun-dational to our moral thought and practices,” and so any attempt to justify one in terms of the other involves us in “a kind of category mistake:” (Magill 2000: 193–94; compare item 1997: ch. 2 and Mackie 1985).[28]

Dennett, as we have already noted, dismisses worries about ultimacy on the ground that they depend on a traditional absolutist conception of responsibility (that is, “guilty-before-the-eyes-of-God”) that is simply unintelligible and should be (moderately?) “revised” in favor of a pragmatic conception based on “moral engineering” by means of sanctions. Against this, Strawson and his followers (for instance, Magill) argue that if compatibilists paid more attention to the role of moral sentiments in this sphere they could provide a richer, nonutilitarian understanding of responsibility. To the extent that this approach remains closer to the original spirit of Dennett's descriptive project, it is more satisfying than the revisionary, pragmatic account of responsibility that Dennett defends. What is not so evident, however, is that the Strawsonian view succeeds in providing us with a sure and easy way of setting aside pessimist worries at the horizon.[29]

It may be argued by the incompatibilist, for example, that our moral sentiments must be targeted only on individuals who possess some relevant set of moral capacities, and that this includes a capacity for ultimate control. Agents who have no control over the specific reason-responsive mechanisms that they have acquired, nor over how these mechanisms are actually exercised in particular circumstances, lack the kind of (ultimate) self-control that is required to sustain and support our moral sentiments. Human beings may possess reason-responsive mechanisms, and be (complex) self-controllers of the kind that Dennett

and others have described, and yet still exercise these capacities in ways that stem ultimately from factors that they cannot control. In some sense, therefore, they have no final say about the moral quality of their own character and conduct.[30] It is not obvious, says the pessimist, that moral sentiments can be sustained when such considerations are pressed upon us.[31]

(p. 249) 6. Pessimism and the Unbearable Limits of Finitude

In my view, the important and significant issues facing the new compatibilism of the kind advanced by Dennett lie primarily with problems of ultimacy at the horizon. The spatial metaphor of distance is helpful in this connection because it indicates that these horizon problems do not immediately present themselves to us in everyday moral life. Close-range and middle-distance issues differ in this respect. In our everyday moral dealings, we ask ourselves whether the conduct we are presented with is a product of the agent's own will, and if so, if the agent is a rationally competent (normal) adult, free from manipulation or coercive pressure. Concerns of this kind are part and parcel of ordinary moral life. Nothing about them is “artificial” or a peculiar product of the Western philosophical tradition.

The situation is not so straightforward at the horizon. Regarding worries about ultimacy, Dennett's general diagnosis of the free will problem seems more plausible. When action is produced by the agent's will, and the agent is clearly capable of rational self-control (that is, reason-responsive), further worries about the ultimate origin or source of the agent's will—in the absence of any worries about manipulation—seem remote from our usual concerns and interests. Worries of this kind seem likely to leave a typical moral audience unmoved. One reason why horizon concerns about ultimacy appear disconnected from ordinary moral life is that, unlike close and middle-distance issues, there is no obvious or decisive way to settle them. That is to say, when we raise questions about *ultimacy*, as distinct from issues of rationality and manipulation, there seems no way to *prove* that an agent was their ultimate source. The skeptic can always challenge such claims by arguing that any appearance of ultimate agency simply reflects our ignorance of the relevant causes at work. We become trapped, consequently, in issues and claims that can never be resolved. Beyond this, the skeptic is also likely to argue that it is not even clear what ultimacy *demands*—so how can we ever *verify* that it is satisfied in a given case? Clearly, general considerations of this kind lend credence to Dennett's claim that horizon problems are the artificial product of (overintellectualized) Western philosophy and theology.

There are, nevertheless, a number of reasons for rejecting this complacent attitude to horizon problems. First, worries of this kind—reaching beyond middle-distance problems of rationality and manipulation—emerge in legal contexts, where the problems are by no means the product of artificial philosophical reflection. On the contrary, lawyers and judges are plainly interested in evidence showing that a person accused of a crime had no control over factors that led to it.[32] Second, and relatedly, our understanding of the influence of genetic endowment (p. 250) and the environment on human conduct and charac-

ter is constantly advancing, and this presses horizon issues on us with increasing force—to refuse to consider them seems mere evasion (compare. Klein 1990: 75 and Greenspan 1993). Most important, it will not do to argue, as Dennett and others have done, that because we are unable to provide a coherent account of how ultimate agency is possible, that we can therefore dismiss worries that agents have no *final* control over their character and conduct. On the contrary, it should be obvious that a convinced skeptic on the subject of "libertarian metaphysics" may draw thoroughly pessimistic conclusions from this (as in the views of the "moral skeptic" or the "hard determinist"). Arguing from the impossibility of ultimate agency to the conclusion that there is no basis for pessimism in the realms of freedom and responsibility is an egregious example of Pollyannaism.

There are interesting structural similarities between pessimism as it relates to the free will problem and the question of human mortality. Consider, for example, Pascal's profoundly pessimistic description of the human condition in the following passage:

> Imagine a number of men in chains, all under sentence of death, some of whom are each day butchered in the sight of the others; those remaining see their own condition in that of their fellows, and looking at each other with grief and despair await their turn. This is the image of the human condition. (Pascal 1966: 165/#434)

The conclusion that Pascal draws from this analogy is that "the only good thing in this life is the hope of another life" (ibid: 157/#427). For our purposes, the interesting thing about this passage is that Pascal uses an "intuition pump" to justify *extreme* pessimism about the human condition. If there is no immortal soul and future state, he suggests, then human life is nothing better than a painful period during which we wait to be executed, along with everyone else.

The obvious reply to all this is that it grossly exaggerates and distorts the limits and miseries of human life. Pascal is guilty of the same sort of abuse of intuition pumps that Dennett objects to in the free will problem. However, while we may grant that Pascal's pessimism is *exaggerated*, it does not follow that all worries about human morality and finitude are without foundation. We may, for example, discredit Pascal's pessimism by pointing out (close and middle-distance) pleasures and sources of happiness that can be found within the span of human life. These show that, typically, our experience of human life does not resemble being chained up and waiting to be executed. At some point, however, those of us who are skeptical about the possibility of immortality must confront the reality of the limits of human existence—the duration of a human life is finite. Such reflections do not impose themselves on us in our everyday concerns, so we are not usually depressed or troubled by them. Nevertheless, to the extent that we (p. 251) have the occasion, opportunity, and temperament to think about such matters, most people will find them sobering or rather melancholy to contemplate.[33] The important point is that we may not share Pascal's extreme pessimism on this subject and yet still appreciate why these reflections on human mortality occasion pessimism of some kind. The reasonable position on this subject, therefore, seems to lie somewhere between Pascalian pessimism and Pollyannaish optimism.

These observations on Pascal's pessimism shed light on both what is right and wrong in Dennett's attempt to discredit incompatibilist pessimism. The incom-patibilist pessimism that Dennett has challenged is essentially Pascalian. It involves analogies and metaphors that are more misleading than illuminating. However, it does not follow from this that reflection on the limits of human agency is not disconcerting or unsettling. On the contrary, when we look beyond the close and middle-distance issues that are the focus of Dennett's attention, we must still confront horizon worries about ultimacy. Even if the worries here are not Pascalian, they provide no basis for Pollyannaish optimism.[34]

What these observations show is that, regarding the free will problem, we must carefully identify the source and quality of our pessimism and note the way they are related. More specifically, it is obvious that the *quality* of our pessimism will vary with the (perceived) *source* of worry. For example, Dennett is surely right to say that if close-range worries were justified (for example, we are in chains), then this would be a “terrible” condition. Much the same is true of middle-distance worries, which would also be “awful.” It is not evident, however, that worries at the horizon have this quality or license an extreme negative emotional response. In the first place, concerns of this kind will vary depending on how lucky/unlucky individuals are with respect to their character and conduct.[35] A person of admirable character may occasion no feeling that her condition is “terrible” or “frightening”—unless, of course, we confuse horizon issues with close and middle-distance pessimism. Even a person whose character and conduct is deplorable cannot be assimilated to the condition of a person who is manipulated or incapable of rational self-control. The sort of pessimism occasioned by a lack of ultimacy must be qualitatively different (that is, reflecting a difference in the source of our concern). An awareness of finitude and contingency, as it relates to the (assumed) impossibility of ultimate agency, licenses a more modest sense of being *disconcerted*, rather than any form of Pascalian despair.[36] In general, it is a mistake to assume that incompatibilist pessimism must take the form of an all-or-nothing, homogeneous, and extreme sense of despair at the thought of the implications of determinism. The alternatives available to both the pessimist and the optimist are surely more subtle and nuanced than this.[37]

(p. 252)

Free will and pessimism by degrees		
Range	*	**Scope of Concern**
Close	3	*Freedom of Action*
	3	Is the agent's conduct regulated by his will?
	3	Are the agent's deliberations and choices futile?
	3	
	3	
Middle	3	*Rational Self-Control*
	3	
	3	Is the agent's will responsive to the available reasons and his true values?
	3	
	3	Is the agent subject to control or manipulation by others?
	3	
	3	
[**** The unstable boundary of ordinary moral life … *****]		
	3	
Horizon	3	*Ultimate Agency*
	3	
	3	Is the agent's character and will ultimately determined by factors that he does not control?
	3	

	3	
	3	Does the agent have a final say about the nature of his character and conduct?
	3	
	3	
	3	
Cosmic	3	*Self-Creation*
	3	
	3	Is the agent an absolute, unconditioned (Godlike) self-creator?
	3	

Notes:

(1.) Dennett's way of associating incompatibilism with “pessimism” is also a prominent feature of Strawson's influential essay “Freedom and Resentment” (Strawson 1962). Although this perspective on the free will debate reflects dominant tendencies in incompatibilist literature, there are some important complications to be noted. For example, the incompatibilist pessimist may well be an “optimist” about the existence of (libertarian) free will. Moreover, some incompatibilists would argue that our everyday beliefs and attitudes concerning freedom and responsibility are not *worth* salvaging, and so they find nothing “frightening” or “awful” about doing without them. As we will see, at times this attitude surfaces in Dennett. For the purposes of this chapter, however, I will work within the pessimist/optimist framework that Dennett (and Strawson) have constructed.

(2.) Dennett's methodology is self-consciously modeled after the ordinary language techniques of predecessors such as Ryle and Wittgenstein (Dennett 1984: 6, 18). (*Elbow Room* is dedicated to the memory of Ryle.)

(3.) Dennett argues that worries about free will are “an almost exclusively Western preoccupation” and that for most people “metaphysical freedom has just not been worth worrying about” (Dennett 1984: 4). Clearly, then, Dennett sees his audience as composed primarily of philosophers, who are victims of their own “induced illusions.”

(4.) “Hobbes, Locke, Hume, Moore, Schlick, Ayer, Stevenson, and a host of others have done what can be done, or ought ever to have been needed, to remove the confusions that can make determinism seem to frustrate freedom” (Davidson 1972).

(5.) There are, in my view, significant problems with the efforts of (empiricist) compatibilists to defend their position on the foundations of a regularity theory of causation. For more on this, see Russell (1988).

(6.) “Fatalism says that my morrow is determined no matter how I struggle. This is of course a superstition. Determinism says that my morrow is determined through my struggle…” (Hobart 1934: 82). For criticism of this doctrine, see Russell (2000).

(7.) The usual point of criticism of classical compatibilism is that freedom of action does not imply freedom of will. Rogers Albritton, drawing on the same distinction, argues that an agent who is unable to act according to his own will (that is, faces “obstacles” of some kind) may nevertheless enjoy “perfect and unconditional” freedom of *will* (Albritton 1985). Indeed, Albritton is skeptical about the very possibility of *unfree* will. Even the addict or compulsive, he claims, lacks only strength of will, which is a different matter. However, Albritton does not discuss “bogeymen” cases of the kind that Dennett describes, and these, in my view, show that his unqualified skepticism concerning the possibility of unfree will is misplaced.

(8.) According to Dennett, our deliberations about our “options” requires only “epistemic openness” (Dennett 1984: 122–23).

(9.) Kane comments on this aspect of Dennett's strategy as follows: “[Dennett] plays the old compatibilist tune in a new key. Just as classical compatibilists distinguish *constraint* from *mere causation*, he says we must distinguish *control* from *mere determination*” (Kane 1996a: 70).

(10.) Kane points out, for example, that children, as they reach maturity, “want an autonomy and dignity that they associate with the power to run their own lives,” even though they “know that their parents are well-intentioned toward them” (Kane 1996a; 69).

(11.) It may be objected that no action that we condemn can be judged as fully the agent's own—since it manifests a failure of normative competence. Benson denies this implication on the ground that “we can sometimes freely do what we believe we should not” (Benson 1987: 480).

(12.) Christman argues that to hold “that freedom is a value only in relation to correct moral norms is to ignore the obvious noninstrumental value of self-mastery itself” (Christman 1991: 358).

(13.) Benson's interesting observations on the social dimension of responsibility, and how it relates to issues of normative competence, lead to further questions about the rele-

vance of *emotional* competence to moral agency. I discuss these matters in more detail in Russell n.d.

(14.) There are, as Fischer and Ravizza point out, difficulties associated with "judgments about mechanism individuation" (Fischer and Ravizza 1998: 40n, 51–2n, 113, 216n; 251n). They offer, however, no "general way of specifying when two kinds of mechanism are the same" and rely, instead, on our "intuitive judgments" about such matters.

(15.) It may be argued that our interest in "moral effort" is closely connected with the question of how an agent actually exercises her rational capacities (that is, how she uses "exercise control"). The exact nature of this relationship is, however, open to a number of different interpretations.

(16.) Dennett claims that in the process of moral development "everyone comes out more or less in the same league"—unless they are "singled out as defective" (Dennett 1984: 96). According to this view, normal adults are all "gifted with powers of deliberation" and "self-control" and at this threshold can be treated as (fully) free and responsible agents (ibid.: 98). However, as indicated, this view leaves large problems unaddressed.

(17.) See esp. Klein (1990: ch. 7), for the details of her effort to (partially) reconcile compatibilist and incompatibilist principles.

(18.) One of Klein's particular concerns is to argue that the U-condition is distinct from incompatibilist worries about "could have done otherwise" (Klein 1990: ch. 2). I will not discuss this aspect of her position.

(19.) Classical compatibilists, of course, insist that worries about the *source* of our moral qualities are misplaced, as this does not change the *value* of the qualities themselves. See, for example, Hobart (1934: 84): "It is the stuff certain people are made of that commands our admiration and affection. Where it came from is another question … Its origin cannot take away its value, and it is its value we are recognizing when we praise."

(20.) This is, of course, a familiar objection to "hierarchical" models of free will, such as Frankfurt (1971). For further discussion of this and related points, see Fischer and Ravizza (1993: 25–33).

(21.) The terminology of "self-forming actions" is from Kane (1996a: esp. ch. 6). Klein is a skeptic about the (empirical) possibility of ultimate agency. Kane (1996a) is a sustained and sophisticated attempt to work out the details of a libertarian metaphysics of this kind.

(22.) For an interesting and important effort to deal with this general problem, see Fischer and Ravizza (1998: 230–36).

(23.) It is arguable that our basic concerns about the way we acquire our reason-responsive mechanisms can be reduced to worries about whether we control the actual exercise of these capacities in particular circumstances. Note, for example, that if we had (ultimate) control over how our reason-responsive mechanism is actually exercised in the con-

text of specific conditions, there seems to be no reason to worry about how the general capacity was acquired (for example, even if it was implanted in some deviant manner).

(24.) The distinction between “descriptive” and “revisionary” metaphysics is introduced and explained in Strawson (1959).

(25.) “My conclusions are neither revolutionary nor pessimistic. They are only moderately revisionary: the common wisdom about our place in the universe is roughly right” (Dennett 1984: 19).

(26.) It is evident that worries about ultimacy return on the account of responsibility for character suggested by Audi. Given that we must be able to “trace” character traits to actions that the agent could *control*, it may be argued that these (“self-forming”) actions must satisfy the U-condition. For a libertarian argument along these lines, see Kane (1996a: 38–40).

(27.) Dennett's views on this subject may be compared and contrasted with Hume's. Hume also holds that a person may be (morally) evaluated for character traits over which he has little or no control. Indeed, he takes the more radical view that this in cludes “natural abilities” (intelligence, imagination, and so on), understood as pleasurable or painful qualities of mind. For a discussion of Hume's views, see Russell (1995: ch. 9).

(28.) Magill does not claim, on this basis, that “there are no grounds for being troubled by the suffering caused by punishment and blame” (Magill 1997: 47). On the contrary, his point is that the “true problem” that we face is “a practical one about opposing strains within our moral framework and conflicting (nonmetaphysical) moral sentiments within ourselves” (ibid.: 49). Regarding this problem, he claims, “there canbe no general resolution of the tension between the principle of well-being and the principle of desert” (ibid.: 52). Nevertheless, “if we keep in mind that it is what we care about, informed by our personal, moral and political feelings and sentiments, that generally informs whether we take the objective or the reactive attitudes, we will not be faced with a helpless dilemma every time we confront decisions about whether to blame or to understand” (ibid.: 52). On the subject of moral sentiment and retributivism see also Russell (1995: ch. 10).

(29.) For the details of this, see Russell (1992).

(30.) There are a number of important complexities here that I cannot pursue. Suffice to note in passing, however, that this way of interpreting what is needed to satisfy ultimacy (that is, the forms of control missing from new compatibilist accounts) may set a standard that some suggested libertarian accounts of ultimacy still fail to meet.

(31.) An illuminating discussion of this problem is presented in Watson (1987b); but compare McKenna (1998b).

(32.) See, for example, Clarence Darrow's classic “hard determinist” defence of Leo pold and Loeb (Darrow 1924). It is significant that Darrow did not argue that his clients did not understand what they were doing or lacked general powers of rational self-control.

On the contrary, his defense is based largely on the (assumed) existence of causes of their character and conduct that were ultimately beyond their control. It is also significant, however, that he refers to several different "bogeymen," which tends to obscure the exact nature of his case.

(33.) "Neither the sun nor death can be looked at steadily" (LaRochfoucauld 1678: #26). Although we generally assume that people have some shared *sensibility* about such matters, variations of response can always be found. This need not imply, however, any kind of intellectual confusion about the relevant considerations or issues involved.

(34.) It may be argued, of course, that the only way to escape from pessimistic worries of this kind, is to embrace libertarian metaphysics, much as some maintain that the only way to escape pessimism about the finitude of human life is to embrace the doctrine of the immortality of the soul.

(35.) Compare, for example, our sense of luck regarding the distribution of other qualities such as beauty or intelligence. It is not obvious that the beautiful or intelligent person will feel any sense of "despair" or "fear" when she contemplates her situation—although the (unfortunate) ugly or stupid person may view things differently.

(36.) Although I believe that reflection on horizon issues of ultimacy generate a sense of disconcertment, my reason is not that it threatens, systematically, to discredit our moral sentiments. On the contrary, when we reflect on considerations about the finitude of human agency, the thought that presses upon us is that who we are, and what we are responsible for to other human beings, depends ultimately on factors that we cannot control. This sobering thought makes us aware of the (uncomfortable) gap between our aspiration to be self-made selves and the evidence that this is an illusion. Such problems concern the relationship between fate (understood in terms of the issue of origination) and responsibility. A plausible compatibilism, I maintain, must acknowledge the legitimacy of concerns about origination and accommodate them by allowing for the possibility that agents who are subject to fate may nevertheless be justifiably held responsible. On this see Russell (2000).

(37.) There is, of course, a considerable amount of room to be found between Pascalian pessimism and Pollyannaish optimism in respect of the issues of determinism and origination. Other (divergent) positions of this general kind in the contemporary literature can be found in, for example Honderich (1993), Pereboom (1995), and Smilansky (2000), as well as Russell (2000). Smilansky's position, which involves the claim that illusion about libertarian free will is desirable and "morally necessary," is described in his essay in this volume (ch. 22).

Paul Russell

Paul Russell is Professor in Philosophy at the University of British Columbia. He has held a number of visiting positions, including Stanford University, the University of Pittsburgh, and the University of North Carolina at Chapel Hill. His published work

includes 'Freedom and Moral Sentiment: Hume's Way of Naturalizing Responsibility' (Oxford University Press, 1995) and 'The Riddle of Hume's Treatise: Skepticism, Naturalism, and Irreligion' (Oxford University Press: 2008). In 2010 he was the Fowler Hamilton Visiting Fellow at Christ Church, Oxford.