

***PESSIMISTS, POLLYANNAS AND THE NEW COMPATIBILISM***

*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*

Paul Russell

University of British Columbia

In Robert Kane, editor,  
*The Oxford Handbook of Free Will* (Oxford University Press, 2002)

The aim of this essay is to describe and assess recent contributions to compatibilist literature not discussed elsewhere in the Handbook. Although the views of several authors will be considered, the main focus of this essay is Daniel Dennett's **Elbow Room**, an important work in the evolution of the "new compatibilism".

## I. 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 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 labelled as "pessimistic" in character. Dennett's fundamental objective in **Elbow Room** is to discredit incompatibilist pessimism and to vindicate a more "optimistic" position (Dennett 1984: 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" (Dennett 1984: 4).<sup>1</sup>

The opening chapter of **Elbow Room** provides a vivid and lively account of how incompatibilist pessimism acquires its psychological grip over us. The worries and anxieties that we have 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 some relevant negative emotional response (Dennett 1984: 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.

It is Dennett's view that the analogies and metaphors concerned "do not in the slightest deserve the respect and influence they typically enjoy" (Dennett 1984:7). 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.<sup>2</sup> If the therapy succeeds, then the free will problem, as traditionally conceived, "dissolves".<sup>3</sup>

A particularly important sub-set 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 sub-divided 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 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 sub-set of bogeymen operate on us in a different way. In these cases the worry is not that our willings fail to guide our conduct, but rather that how 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 being hypnotized, or manipulated by an evil neurosurgeon who uses 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 stimulus 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 -- there is no real agent 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 attention on his examples of bogeymen and the two different ways that they pose a threat to human freedom. In order to distinguish 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 and/or we are subject to manipulation of some kind. Finally, I will also consider worries that our will is ultimately determined by causal antecedents that we have no control over. I refer to this concern as pessimism at the horizon.

## **II. Classical Compatibilism and Close-Range Pessimism**

A number of the basic arguments employed by Dennett 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.<sup>4</sup> The classical arguments deal primarily with close-range pessimism. The position taken is that the traditional free will debate is a "pseudo-problem", since it is 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. In 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 of the

agent's is to say only that events of the first kind regularly follow events of the second kind -- nothing more is involved.<sup>5</sup>

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 is simply to confuse two distinct issues.<sup>6</sup> 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).

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 (Hume 1748: 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 1654: I,60-1). The only freedom that we need or want, according to this view, is that we are 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 to the absurdity of an infinite regress.

### **III. 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 well short of dealing with middle-distance pessimism. The most obvious difficulty facing any conception of moral freedom that identifies it with the ability to act according to the determination of an agent's desires or willings is that freedom of this kind 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.<sup>7</sup>

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 behaviour, but rather that we are not able to regulate our will according to reason and/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 how we judge the prospects of the "new compatibilism".

Dennett's interpretation of these cases, and the fears that they generate, centres on two

closely related issues. The first concerns the worry that we are not able to regulate our will in light of the reasons that are present 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. In 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). The reasons that we have 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" (Dennett 1984: 37).<sup>8</sup>

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 and desires will detect "abnormalities" in their causation (Dennett 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 abilities of this kind. 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 (Dennett 1984: 12,34,37-8).

Central to getting a clear idea of 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" (Dennett 1984: 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 (Dennett 1984: 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 (i.e. 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 (Dennett 1984: 62-3).<sup>9</sup> 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 control.<sup>10</sup>

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 (Dennett 1984: 133-5). Beyond this, if such a condition had to be satisfied to establish responsibility -- i.e. the agent could have done otherwise in 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 useless. What we want to know is if 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 in the specific circumstances she acted in. The question that matters to us is whether or not there is a flawed character trait that needs to be corrected (Dennett 1984: 137-8).

According to Dennett there is another insidious (middle-distance) bugbear that needs to be exorcised from the overactive incompatibilist imagination -- one that is, he believes, especially powerful in its hold over us. This is the worry that if determinism is true then we are (somehow) "controlled by nature" and/or "controlled by the past" (Dennett 1984: 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 (Hume 1748: 99-103). Clearly God is not conceived of as evil, but in respect of 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 to this is that these 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 be in control of it (Dennett 1984: 72). All talk of being controlled by Nature or by the Past plainly involves personification (Dennett 1984: 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 the accompanying resistance to the thesis of determinism.<sup>11</sup>

Dennett's analysis of these incompatibilist worries covers three issues that need to be carefully separated. They are: (1) Do human purposes and choices have determining causes that ultimately originate outside of them?; (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 (e.g. "hideous hypnotist" etc.). 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".<sup>12</sup> 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.

Dennett associates worries about the ultimate origin of our deliberations and choices with the aspiration to "absolute agenthood" -- to be a perfect, God-like self-creator (Dennett 1984: 83-5). 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 (i.e. "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 serve as 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.

#### **IV. 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 papers by Paul Benson. In "Freedom and Value" Benson argues that free agency requires another "equally significant ability" apart from control. This is 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 competent 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" (Benson 1987: 472).

Benson points out that the omission that he is concerned with in compatibilist accounts of moral capacity (i.e. normative competence) is addressed in Susan Wolf's paper "Asymmetrical Freedom" (Wolf 1980; and see also Wolf 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). Since the norms governing our actions are important to us, so too must the ability to regulate our conduct by means of our evaluative judgments. Benson continues:

...if 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" (Benson 1987: 475).

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" (Benson 1987: 482). When we do not have any control over what we do (e.g. 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 (e.g. 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 (e.g. older children, the severely deprived etc.).<sup>13</sup> 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".<sup>14</sup>

In a more recent paper "Free Agency and Self-Worth" Benson modifies his position. He argues, in this context, that the "normative-competence condition" is too strong, in so far 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" (Benson 1994: 665). 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).

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. These kinds of condition 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" (Benson 1994: 661). Related to this point, this condition of free agency also makes clear 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" (Benson 1994: 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" (Benson 1994: 668).<sup>15</sup>

Another important set of issues that arises from Dennett's discussion is 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. Their position is discussed in detail elsewhere in the Handbook (see Ish Haji's article), but a few brief comments are in order. 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" (Fischer & Ravizza 1998: esp. 34-41). A free agent, on a "strong" interpretation, operates with a mechanism that is always receptive and responsive to available reasons. In these circumstances, the agent's reasons, choices and actions all successfully "track value" or "the reasons there are" (Fischer & Ravizza 1998: 42). 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, that can serve the purposes of compatibilism, must describe "mechanisms" that can fail under some conditions, without being systematically unreliable (i.e. too "weak"). We need, therefore, some principled way of distinguishing and identifying sufficiently reliable mechanisms. This problem leads on to another major difficulty for accounts of this kind. 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 the way that the mechanism or capacity is actually exercised in the particular conditions of action. This deep problem is closely connected with the issue of "ultimacy" or "absolute agency", and I will return to it when I discuss pessimism at the horizon.

It is not only the possibility that mechanisms may fail to respond to reasons that present difficulties, there are also difficulties associated with "strong" mechanisms that cannot fail (i.e. 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: 130-3). 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; cp. Wolf 1990: 138-42).

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 a straightforward matter. 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.<sup>16</sup>

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.<sup>17</sup> Dennett's strategy is to argue that our (natural) complexity -- not indeterminism -- provides us with the ability to be sensitive 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 any more.

## **V. 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 be responsive to reason and we do not want another agent controlling 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: 100,156-7). 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

vissitudes 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" (Dennett 1984: 72-3).

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 (Dennett 1984: 33,75,83-4). He cites, for example, A.J. Ayer's description of "implanted" desires and beliefs (Ayer 1980: 9); Paul Edwards' 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 want to turn to Martha Klein's particular account of this problem.

What is fundamental to the incompatibilist position, Klein argues, is the conviction "that one of the things which disqualifies an agent from blameworthiness is his not having been responsible for the causes of his decisions or choices" (Klein 1990: 51). This conviction commits us, she says, to a "U-condition" for agent accountability. This is 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" (Klein 1990: 51).<sup>18</sup> Klein takes the basic rationale behind the U-condition to be 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 (Klein 1990: 50).<sup>19</sup>

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 tumours, implantations, brainwashing and hypnosis. What these cases have in common, Klein maintains, is that the agent's decisions can be traced to causes for which he is not responsible, and so he ought not to be blamed (Klein 1990:70). The example of the brain tumour is especially important to Klein's case for the U-condition because it highlights the point that it is not worries about the "implantation" of desires and beliefs by others, as such, that is the real source of worry, 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" (Klein 1990: 73).

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" (Klein 1990: 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 (Klein 1990: 165-6). In 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 will be that his observations on middle-distance pessimism, and the bogeymen that it conjures up, serve to 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: 33-4). 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 worrying about brain tumours 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 (cp. Dennett 1984: 64-5). While this is something that is

frightening, there is no basis for supposing that it implied by determinism. 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 it is true that our beliefs and desires may be subject to self-monitoring activities of various kinds, the fact remains that these activities are themselves conditioned by factors that are not of the agent's own making.<sup>20</sup> 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.<sup>21</sup> The sorts of capacities that Dennett and other new compatibilists in his mould describe fall short of this, and so their strategy does not to relieve the pessimist of worries at the horizon.

There are, however, further avenues of reply 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 considerers" (Dennett 1984: 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 that we have no final control over (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.<sup>22</sup> (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. If determinism holds sway, then how capacities for self-creation and self-monitoring are exercised in a given situation will ultimately be determined by factors that the agent has no control over.<sup>23</sup> Dennett is clearly right to say that 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" (Dennett 1984: 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".<sup>23</sup> That is to say, the position taken 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".<sup>25</sup> The argument here is that worries about ultimacy may be motivated by a conception of responsibility that, although historically 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" (Dennett 1984:165-6).<sup>26</sup> 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 in character, and he leans heavily on "engineering" metaphors when describing how this system operates (Dennett 1984: Chps. 6,7). Responsibility, he argues, should be understood in terms of "the rationale of punishment", and its rationale is that it supports the criminal laws of society.<sup>27</sup> That is to say, we punish individuals in those cases where we think they are "mentally competent" enough to be deterred or reformed by the threat or imposition of sanctions. All this is not only highly "revisionary" in approach, it also takes a (large) step back in the direction of classical compatibilism.<sup>28</sup>

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). A central theme of Strawson's paper 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 to "reactive attitudes and feelings" or "moral sentiments". These 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: 78). To a limited extent we can suppress these reactions in particular cases or circumstances, however, there is no question of us being able systematically to abandon or suspend our commitment to the whole "complicated web of attitudes and feelings".<sup>29</sup>

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. The pessimist is right in saying that a purely utilitarian approach to responsibility leaves out "something vital in our conception of these practices" (Strawson 1962: 78). On this basis, however, they mistakenly conclude that what is required to fill this "lacuna" in the optimist's account is some form of libertarian metaphysics that involves denying determinism (Strawson 1962: 78-80). 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 (Strawson 1962: 74; cp.68,70). To suppose otherwise is "to over-intellectualize the facts" (Strawson 1962: 78). When the role of moral sentiment is allowed its proper place in moral life, we can avoid both a crude utilitarian account of responsibility which is divorced from psychological reality, while at the same time steering clear of 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 an essential feature of human life as we know it.

Dennett dismisses worries about ultimacy on the ground that they depend on a traditional absolutist conception of responsibility (i.e. "guilty-before-the-eyes-of-God") that is simply unintelligible, and should be (moderately?) "revised" in favour of a pragmatic conception that relies on "moral engineering" by means of sanctions. Against this, Strawson and his followers argue that if compatibilists paid more attention to the role of moral sentiments in this sphere they could provide a richer, non-utilitarian understanding of responsibility. It is certainly true that this approach is closer to the original spirit of Dennett's descriptive project and to this extent it is more satisfying, since it remains committed to our established attitudes and practices. However, whatever the merits of this approach, I am not as confident as Strawson and some of his followers that an understanding of the role of moral sentiment provides, by itself, a sure or easy way of setting aside pessimist worries at the horizon.<sup>30</sup>

It remains open to the incompatibilist to argue 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 has described, and yet still exercise these capacities in ways that are ultimately due to factors that they have no control over. In some sense, therefore, they have no final say about the moral quality of their own character and conduct.<sup>31</sup> It is not obvious, says the pessimist, that moral sentiments can be sustained when considerations of this kind are pressed upon us.

## **VI. 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 are not something that 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 if 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. There is nothing about them that is "artificial" or a peculiar product of the Western philosophical tradition.

The situation is not so straight-forward at the horizon. In respect of 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 (i.e. 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 sceptic can always challenge claims of this kind 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 sceptic 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 give credence to Dennett's claim that horizon problems are the artificial product of (over-intellectualized) 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 at issue 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.<sup>32</sup> Second, and relatedly, our understanding of the influence of genetic endowment and the environment on human conduct and character is constantly advancing, and this presses horizon issues on us with increasing force -- to refuse to consider them seems mere evasion (Cp. Klein 1990: 75). Finally, and most importantly, 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 sceptic about

"libertarian metaphysics" may draw thoroughly pessimistic conclusions from this (e.g. as in the views of the "moral sceptic" or the "hard determinist"). Arguing from the impossibility of ultimate agency to the conclusion that there is no basis for pessimism in respect of freedom and responsibility is an egregious example of pollyannism.

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" (Pascal 1966: 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 is a gross exaggeration and distortion of 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 there is nothing to be said for worries about human mortality or finitude on the horizon. We may set about, for example, to 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 is not like being chained up and waiting to be executed. At some point, however, those of us who are sceptical about the possibility of immortality, must confront the reality of the limits of human existence -- the duration of a human life is finite. Reflections of this kind 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 have the occasion, opportunity and temperament to think about such matters, most people will find them sobering or rather melancholy to contemplate.<sup>33</sup> 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 pollyannish optimism.

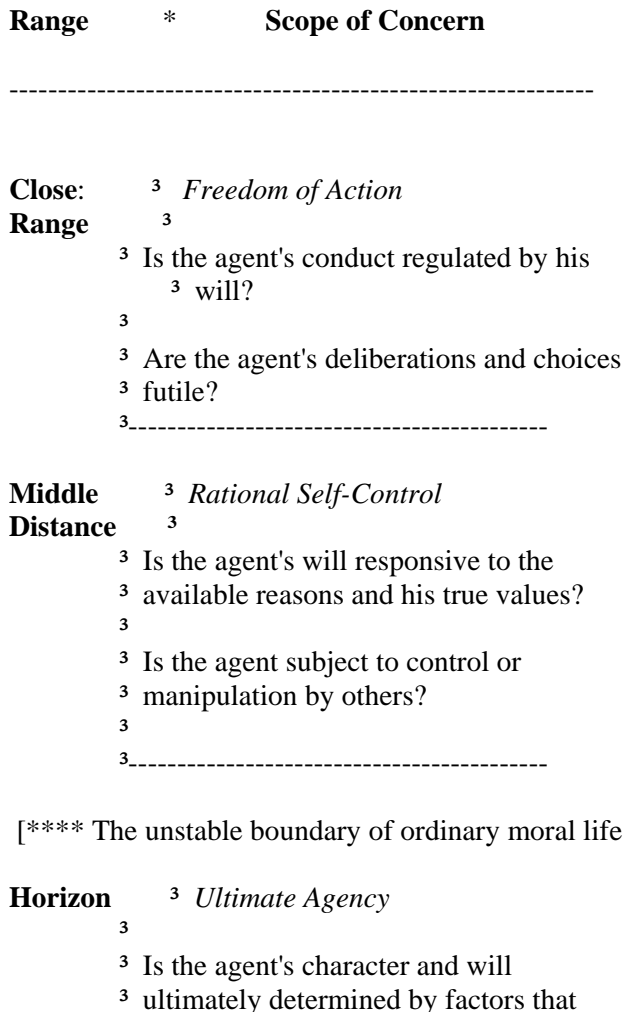
These observations on Pascal's pessimism shed light on both what is right and wrong in Dennett's attempt to discredit incompatibilist pessimism. The incompatibilist pessimism that Dennett has challenge is essentially Pascalian in character. It involves analogies and metaphors that are more misleading than illuminating. However, it does not follow from this that reflection on the limits of human agency provide us with no basis for feeling disconcerted or unsettled. On the contrary, when we look beyond the close and middle-distance issues that Dennett has focused on, we must still confront horizon worries about ultimacy. Even if the worries here are not Pascalian in character, they are no basis for pollyannish optimism.<sup>34</sup>

What these observations show is that in respect of the free will problem we must carefully discriminate the source and quality of our pessimism, and take note of the way they are related. More specifically, it is obvious that the quality of the pessimism that we experience will vary with the (perceived) source of worry. For example, Dennett is surely right to say that if close-range worries were justified (e.g. 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.<sup>35</sup> 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 that a lack of ultimacy occasions must be qualitatively different (i.e. 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.<sup>36</sup> 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.

**Diagram:**

**Free Will and Pessimism by Degrees**



- <sup>3</sup> he does not control?
- <sup>3</sup>
- <sup>3</sup> Does the agent have a final say about
- <sup>3</sup> the nature of his character and conduct?
- <sup>3</sup>
- <sup>3</sup>-----

**Cosmic**      <sup>3</sup> *Self-Creation*

- <sup>3</sup>
- <sup>3</sup> Is the agent an absolute, unconditioned
- <sup>3</sup> (God-like) self-creator?
- <sup>3</sup>

## FOOTNOTES

- (1) Dennett's way of associating incompatibilism with "pessimism" is also a prominent feature of Strawson's influential paper "Freedom and Resentment" (Strawson 1962). Although this way of looking at 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 ("revisionary") attitude surfaces in Dennett. For the purposes of this essay, however, I will work within the pessimist/optimist framework that Dennett (and Strawson) have constructed.
- (2) The methodology that Dennett employs is self-consciously modelled after the ordinary language techniques of predecessors such as Ryle and Wittgenstein (Dennett 1984: 6,18).
- (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 comprised primarily of philosophers, who are victims of their own "induced illusions".
- (4) Classical compatibilism still has, of course, distinguished defenders. For example, Donald Davidson remarks: "... 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 1980: 63).
- (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). On this doctrine see Russell 2000.
- (7) The usual point of criticism made against 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 (i.e. faces "obstacles" of some kind), may nevertheless enjoy "perfect and unconditional" freedom of will. Indeed, Albritton is sceptical 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 scepticism concerning possibility unfree will is misplaced.

- (8) Dennett's account of our reflective capacity is, of course, closely related to other "hierarchical" or "real self" theories of freedom, as advanced by, e.g., Frankfurt 1971 and Watson 1975.
- (9) According to Dennett, our deliberations about our "options" requires only "epistemic openness" (Dennett 1984: 122-3).
- (10) One implication of this understanding of control is that there are "degrees of freedom" (Dennett 1984: 53). See also Williams 1985: 5.
- (11) 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 1996: 70).
- (12) 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 1996: 69).
- (13) For a rather different compatibilist perspective on the issue of deprivation and blameworthiness see Klein 1990: esp. Chp. 4, sec.3. For another view, closer to Benson's, see Wallace 1994: 231-5.
- (14) 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).
- (15) Benson's interesting observations on the social dimension of responsibility, and how it relates to issues of normative competence, lead on to further questions about the relevance of emotional competence to moral agency. I discuss these matters in more detail in Russell N.d.
- (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 (Dennett 1984: 98). However, as indicated, this view leaves large problems unaddressed.
- (17) See also Wallace 1994: 175-7, for a related account of how "problem cases" of these kinds can be interpreted in terms of a break-down of rational self-control.
- (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: Chp.2) I will not discuss this aspect of her position.
- (19) Klein's specific way of interpreting the U-condition and its significance is open to revision. See, in particular, Kane 1996: esp. Chps. 3 and 5.
- (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 Russell 1995: 128-33.
- (21) The terminology of "self-forming actions" is from Kane 1996: esp. Chp.6. Klein is a sceptic about the (empirical) possibility of ultimate agency. Kane's 1996 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 how we acquire our reason-responsive mechanisms, can be reduced to worries about whether we control how these capacities are actually exercised in particular circumstances. Note, for example, that if we had (ultimate) control over how our reason-responsive mechanism is actually exercised in specific

- conditions, there seems no reason to worry about how it was acquired (e.g. 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) On related themes see Williams 1985.
- (27) There are interesting features of Dennett's conception of responsibility that are get obscured by his utilitarian-pragmatic orientation. For example, Dennett discusses the importance of character traits -- as distinct from individual actions -- for our understanding and interest in issues of responsibility. In general, the relevance of character to questions of responsibility (for action) has received insufficient attention in the free will debate. A notable exception to this is Robert Audi, who argues that "all (normative) responsibility traces to acts and ultimately to basic acts" (Audi 1991: 307). Audi uses this principle to suggest a compatibilist account of responsibility for character. Audi's position is worth contrasting with Hume's, which is examined in Russell 1995, Chp.9.
- (28) One of the ironies about Dennett's pragmatic-engineering approach to responsibility is that it lets real worries about manipulation and "conditioning" to resurface. There is, in fact, something of the spirit of B.F. Skinner's *Walden Two* to be found in Dennett's views on this subject. A good account of why we should be troubled by this is found in Kane 1996: 65-70, 201-04.
- (29) Strawson's (Humean) naturalistic strategy has been widely discussed and debated, and a number of sympathetic studies have appeared that suggest various modifications and/or additions to it. Several of these neo-Strawsonian works are discussed elsewhere in the Handbook. (See Ish Haji's article.) In this context, however, particular note should be taken of recent work by Kevin Magill, since his defence of Strawson is directly relevant to Dennett's utilitarian approach to questions of responsibility. Picking up on themes from Strawson, Magill argues 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 (i.e. that the guilty should suffer). According to Magill, both the utilitarian and retributive principles are "foundational 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: 193-4; cp. Magill 1997: 44-6).
- (30) For the details of this see Russell 1992.
- (31) There are a number of important complexities here that I cannot pursue. Suffice to note in passing, however, that if we interpret ultimacy in these terms (i.e. with a view to the forms of control missing from new compatibilist accounts) then we set a standard that some suggested libertarian accounts of ultimacy will fail to meet.
- (32) See, e.g., Clarence Darrow's classic "hard determinist" defence of Leopold 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 defence is based largely on the (assumed) existence of ultimate causes of their character and conduct, which were 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). As the example makes clear, we generally assume that people have some shared sensibility about such matters. Variations of response can, of course, always be found, but this need not imply 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 in respect of 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 serve to generate a sense of being disconcerted, I do not believe this because 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 is pressed upon us is that who we are, and what we are responsible for to other human beings, depends ultimately on factors that we have no control over. This is certainly a sobering thought, since it makes us aware of the (uncomfortable) gap between our aspiration to be "self-made-selves" and the evidence that this is an illusion. These are issues that concern the relationship between "fate" and responsibility. (However, the meaning of "fate" in this context is very different from the classical compatibilist understanding of the term.) 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.

## Bibliography

- Albritton, Rogers. 1985. "Freedom of Will and Freedom of Action." **Proceedings and Addresses of the American Philosophical Association** 59: 239-51.
- Audi, Robert. 1991. "Responsible Action and Virtuous Character." *Ethics* 101: 304-21.
- Ayer, A.J. 1980. "Free-will and Rationality", in Z. van Straaten (ed.), 1980.
- Benson, Paul. 1987. "Freedom and Value." **Journal of Philosophy** 84: 465-86.
- Benson, Paul. 1994. "Free Agency and Self-Worth." **Journal of Philosophy** 91: 650-68.
- Berofsky, Bernard (ed.). 1966. **Free Will and Determinism**. New York and London: Harper and Row.
- Campbell, C.A. 1951. "Is 'Freewill' a Pseudo-Problem?", reprinted in B. Berofsky (ed.), 1966.
- Darrow, Clarence. 1924. "The Crime of Compulsion". Reprinted in A. Weinberg (ed.), Clarence Darrow, **Attorney for the Damned**. New York: Simon and Schuster, 1957.
- Davidson, Donald. 1973. "Freedom to Act", reprinted in **Essays on Actions and Events**. Oxford: Clarendon, 1980.
- Dennett, Daniel. 1984. **Elbow Room**. Oxford: Clarendon Press.
- Edwards, Paul. 1961. "Hard and Soft Determinism", in S. Hook (ed.), 1961.
- Fischer, John (ed.). 1986. **Moral Responsibility**. Ithaca and London: Cornell University Press.
- Fischer, John and M. Ravizza (eds.). 1993. **Perspectives on Moral Responsibility**. Ithaca and London: Cornell University Press, 1993.
- Fischer, John, and M. Ravizza. 1998. **Responsibility and Control: A Theory of Moral Responsibility**. Cambridge: Cambridge University Press.
- Frankfurt, Harry. 1969. "Alternate Possibilities and Moral Responsibility", reprinted in J. Fischer (ed.), 1986.
- Frankfurt, Harry. 1971. "Freedom of the Will and the Concept of a Person", reprinted in G. Watson (ed.), 1982.
- Hobart, R.E. 1934. "Free Will as Involving Determination and Inconceivable Without It", reprinted in B. Berofsky (ed.), 1966.
- Hobbes, Thomas. 1654. **Of Liberty and Necessity**, selections reprinted in D.D. Raphael (ed.), **British Moralists 1650-1800**. 2 Vols. Oxford: Clarendon Press, 1969.
- Hook, Sidney (ed.). 1961. **Determinism and Freedom in the Age of Modern Science**. New York: Collier, 1961.
- Hume, David. 1748. **Enquiries Concerning Human Understanding and Concerning the Principles of Morals**. L.A. Selby-Bigge, editor., 3rd rev. ed. by P.H. Nidditch. Oxford: Clarendon Press, 1975.
- Kane, Robert. 1996. **The Significance of Free Will**. New York and Oxford: Oxford University Press.
- Klein, Martha. 1990. **Determinism, Blameworthiness, and Deprivation**. Oxford: Clarendon Press.
- LaRochfoucauld, Francois. 1678. **Maxims**. Translated by L. Tancock. Harmondsworth, Middx.: Penguin, 1959.
- Magill, Kevin. 1997. **Freedom and Experience: Self-Determination without Illusions**. London, Macmillan.
- Magill, Kevin. 2000. "Blaming, Understanding and Justification", in T. van den Beld (ed.), 2000.
- Nagel, Thomas. 1976. "Moral Luck", reprinted in G. Watson (ed.), 1982.
- Pascal, Blaise. 1666. **Pensees**. Translated and introduced by A.J. Krailsheimer. Harmondsworth, Middx.: Penguin.
- Russell, Paul. 1988. "Causation, Compulsion and Compatibilism." **American Philosophical Quarterly** 25: 313-21.
- Russell, Paul. 1992. "Strawson's Way of Naturalizing Responsibility." *Ethics* 102: 287-302.
- Russell, Paul. 1995. **Freedom and Moral Sentiment: Hume's Way of Naturalizing**

- Responsibility.** New York and Oxford: Oxford University Press.
- Russell, Paul. 2000. "Compatibilist-Fatalism", in T. van den Beld (ed.), 2000.
- Russell, Paul. N.d. "Responsibility, Moral Sense and Symmetry".
- Schlick, Moritz. 1939. "When is a Man Responsible?", reprinted in B. Berofsky (ed.), 1966.
- Skinner, B.F. 1971. **Beyond Freedom and Dignity.** New York: Vintage Books.
- Strawson, P.F. 1959. **Individuals: An Essay in Descriptive Metaphysics.** London: Methuen.
- Strawson, P.F. 1962. "Freedom and Resentment", reprinted in G. Watson (ed.), *Free Will*, 1982.
- van den Beld, Ton (ed.). 2000. **Moral Responsibility and Ontology.** Dordrecht: Kluwer.
- van Straaten, Zak (ed.). 1980. **Philosophical Subjects: Essays Presented to P.F. Strawson.** Oxford: Oxford University Press.
- Wallace, R. Jay. 1994. **Responsibility and the Moral Sentiments.** Cambridge, Mass.: Harvard University Press.
- Watson, Gary. 1975. "Free Agency", reprinted in G. Watson (ed.), 1982.
- Watson, Gary (ed.). 1982. **Free Will.** Oxford University Press.
- Watson, Gary. 1987. "Responsibility and the Limits of Evil", reprinted in J. Fischer and M. Ravizza (eds.), 1993.
- Watson, Gary. 1986. Review of Dennett's *Elbow Room*. **Journal of Philosophy** 83: 517-22.
- Williams, Bernard. 1985. "How free does the will need to be?", in **Making sense of humanity.** Cambridge: Cambridge University Press, 1995.
- Wolf, Susan. 1980. "Asymmetrical Freedom." **Journal of Philosophy** 77: 151-60.
- Wolf, Susan. 1990. **Freedom Within Reason.** New York and Oxford: Oxford University Press.\_