***Moral Responsibility and Existential Attitudes***

Paul Russell

Lund University/University of British Columbia

August 2018/ Revised June 2020

[Forthcoming in D. Nelkin & D. Pereboom, eds,

*Oxford Handbook of Moral Responsibility*, Oxford University Press.]

*Philosophy, and in particular moral philosophy, is still deeply attached to giving good news.*

- Bernard Williams

**I. Existential Attitudes and the Human Predicament**

Our understanding and interpretation of human agency and moral responsibility is evidently not just a matter of theoretical interest and curiosity. It affects and modifies our attitude to our lives and how we may lead them. In this respect the philosophical issue of human freedom and moral responsibility is similar to the problems of the existence of God and the immortality of the soul. These are, as Kant noted, issues that concern not just what we know but what we ought to do and what we may hope for.[[1]](#endnote-1) We *care* about these matters because they serve to guide and inform how we lead our lives and why we may want to continue them. With respect to these matters, therefore, it is understandable that we may seek some measure or degree of philosophical consolation and reassurance. It is, nevertheless, entirely possible the truth about our predicament will disappoint these hopes and aspirations.

We may refer to metaphysical problems that carry significance for the human predicament as *existential* metaphysical problems. In confronting problems of this sort we seek to provide some sort of vindication of our lives and existence – both as individuals and collectively. The challenge is “to make sense of human life” within a credible world-view. These problems share several important features. First, they generally arouse attitudes that are either positive or negative in *quality*. They are positive insofar as they affirm or sustain some feature or aspect of our existence that we value or care about. Conversely they are negative when they have the opposite significance. We might refer to this as the *valence* of our existential attitudes. Second, this valence may also vary as a matter of *degree* – being more or less intense. The intensity of a given attitude will vary with either the weight and significance of what it is that we attach value to or with the extent to which it is perceived to be threatened (e.g. in light of skeptical arguments). Finally, when we consider existential metaphysical issues of this kind, we may distinguish between an *orthodox* and *unorthodox* interpretation of the relationship between skepticism and pessimism. Generally speaking, we assume that skepticism about the feature in question justifies some measure or degree of pessimism, on the ground that the feature in question is required to secure something of particular value or worth for us. It is clear, nevertheless, that skeptical conclusions do not always generate pessimism, as some may accept a given skeptical conclusion but go on to *deny* that this licences any form of pessimism. As we shall see, when we consider the issue of moral responsibility, the general structure of our available options falls into this general structure or pattern, as found in various other existential metaphysical problems.

**II. Strawson’s Taxonomy**

Our primary task, in this chapter, is to provide an overview of the available positions relating to existential attitudes and moral responsibility. For this purpose, as it concerns the contemporary debate, the right place to begin is with P.F. Strawson’s enormously influential discussion in “Freedom and Resentment” (first published in 1962). In this paper Strawson introduces the labels of “optimism” and “pessimism” as a way of categorizing the opposing parties in the debate concerning the relationship between moral responsibility and the thesis of determinism. He describes the “Pessimists” as those who hold that *if* the thesis of determinism is true, then the concepts of moral obligation and responsibility really have no application, and the practices of punishing and blaming, of expressing moral condemnation and approval are really unjustified (Strawson 1962/2013: 63). The “Optimists”, however, “hold that these concepts and practices in no way lose their *raison d'être* if the thesis of determinism is true” (Strawson 1962/2013: 63). (Hereafter I will use capital letters when citing the specific labels that Strawson employs.) Strawson also distinguishes a third party in this debate, “the genuine moral skeptic”, who takes the view “that the notions of moral guilt, of blame, of moral responsibility are inherently confused and that we can see this to be so if we consider the consequences of either the truth of determinism or its falsity” (Strawson 1962/2013: 63).

Strawson’s tripartite analysis is in some respects confusing and problematic. First, there is the question concerning the relationship between the Pessimist and the Skeptic. The Pessimist, as Strawson usually presents this position, is a *libertarian* Pessimist, who believes that determinism is false *and* that we do have free will. Both the libertarian Pessimist and the Skeptic are *incompatibilists*. The Skeptic, however, denies that we have free will, whether determinism is true or false. The Skeptic, therefore, rejects libertarian metaphysics (e.g. “contra-causal freedom” of some kind) and maintains that there are *no free and responsible moral agents* – which is certainly a dramatic conclusion. Since the libertarian Pessimist is *not* a Skeptic – and believes that humans are moral agents who have free will – there is an important sense in which the libertarian Pessimist is actually *optimistic*. Clearly, then, both the libertarian Pessimist and the (compatibilist) Optimist share the (optimistic) view that we enjoy free will of the relevant kind that supports moral responsibility. Where they differ, is that the Optimist believes that this does not require the falsity of determinism, whereas the libertarian Pessimist believes that it does.

Where, then, Strawson stands in relation to the Optimist/Pessimist divide? Contrary to the Pessimist, Strawson rejects the suggestion that the truth of determinism implies skepticism about moral responsibility. Strawson’s argument for this is complex and not all of its details need concern us here but the crucial element he relies on to defeat skepticism is that of the “reactive attitudes” or “moral sentiments”, along with a theory of excuses and exemptions. According to Strawson’s account, the foundations of responsibility rest with the fact that we *care* about the attitudes and intentions of other agents and the extent to which they do or do not manifest “some degree of good will or regard” (Strawson, 1962/2013: 66-7). Reactive attitudes are, it is argued, “natural human reactions to the good or ill will or indifference of others towards us, as displayed in *their* attitudes and actions” (Strawson, 1962/2013: 70 – Strawson’s emphasis). Our proneness to such reactions is exactly what is involved in making these demands – such that when they are violated we naturally respond with feelings of indignation, resentment and blame. It is, according to Strawson, these fundamental features of human moral psychology that are essential if we are to make sense of the conditions of moral responsibility and the kind of justification that it may require (Strawson, 1962/2013: 64).

Understanding responsibility in terms of the reactive attitudes serves to explain why skepticism of this general kind would be both *unliveable* and *unbearable* in practice. Strawson’s key arguments against skepticism turn on the rationale of excuses and exemptions. Nothing about the (assumed) truth of determinism implies that our reactive attitudes are never justified. In respect of excuses, such as ignorance, accidents or physical force, none of these considerations apply generally simply because determinism is true. Similarly, in the case of exemptions, where we regard a person as an inappropriate target of reactive attitudes, there is nothing about determinism that suggests that all agents are somehow insane, immature or severely disabled. In circumstances of these kinds, where agents are morally incapacitated, we are required to systematically withdraw our reactive attitudes. This involves dropping the participant stance and adopting a sustained attitude of “objectivity” (Strawson, 1962/2013: 69-72). An attitude of this kind, Strawson maintains, requires us to view the agent “as excluded from ordinary adult human relationships” and as an individual who is to be “managed or handled or cured or trained” – someone we may regard in a more detached manner and as not belonging to “the moral community” (Strawson, 1962/2013: 76, 79, 80). If the truth of the thesis of determinism implied that we should always, universally, adopt “a sustained objectivity of inter-personal attitude” that would, indeed, be deeply disturbing. However, not only is this not required of us, any attempt to *systematically* adopt the objective attitude in this way is psychologically impossible for human beings (Strawson, 1962/2013: 71-3).

Let us now turn to Strawson’s particular understanding of the relationship between skepticism and pessimism.[[2]](#endnote-2) Although Strawson believes that the relevant source of pessimism rests with the threat of skepticism, and that skepticism can be discredited, it is also clear that he distances himself from Optimism (i.e. classical compatibilism) in crucial respects. The Optimist, as Strawson describes this position, places emphasis on “the efficacy of our practices of punishment, and of moral condemnation and approval, in regulating behaviour in socially desirable ways” (Strawson, 1962/2013: 64; see also 65, 78-80). This pragmatic, utilitarian perspective is not, Strawson maintains, a sufficient basis or even *the right sort of basis* on which to regulate and justify these attitudes and practices (Strawson, 1962/2013: 65). What it leaves out is the “vital” role of the reactive attitudes. The Pessimist, Strawson argues, is *correct* in claiming that there is a “lacuna in the Optimist’s story” but mistakenly tries and plug this gap with the extravagant and implausible metaphysics of libertarianism. In this respect, both Optimists and Pessimists are guilty of the shared mistake of “over-intellectualizing the facts” and overlooking or ignoring the crucial role of reactive attitudes or moral sentiments in describing the *real* basis of these problematic concepts and practices.

Viewed this way, Strawson’s basic objection to the Optimist’s position is that the stance they adopt is in important respects like that of the Skeptic – viewing agents in a detached manner and as mere objects for treatment, control and manipulation. Strawson agrees with the Pessimist that the Optimist outlook, with its “one-eyed utilitarian” emphasis on controlling future behaviour, is dehumanizing and to a considerable extent shares the unattractive features of the Skeptic’s outlook (Strawson, 1962/2013: 80). From the perspective of Strawson’s critique an exclusively objective stance is not a basis for optimism of any kind but rather a bleak and humanly impoverished form of existence. Even if we *could* embrace it – and Strawson questions this assumption – we have every reason to *resist* it (Strawson, 1962/2013: 72-3). In this way, Strawson’s *modified* optimism rests with his effort to *preserve* a central role for reactive attitudes in human moral life, as opposed to a detached stance of controlling and manipulating agents using the techniques of (behaviourial) science.[[3]](#endnote-3)

There are, as already noted, many possible objections and criticisms that could be levelled against Strawson’s own brand of (compatibilist-friendly) optimism. Two criticisms are, however, of particular interest for this discussion. The first is that Strawson both *exaggerates* the pessimistic implications of skepticism and, at the same time, *underestimates* the significant sources of pessimism that persist for his own account. With respect to the first of these, Strawson maintains that if we were to accept the case for skepticism, and cease entirely to entertain reactive attitudes, this would result in the universal adoption of the objective attitude. A world that operated exclusively on the basis of the objective attitude would, he claims, be a bleak, inhuman world, stripped of all personal relations and emotional ties. Strawson’s critics have argued that his understanding of the reactive attitudes, and their relation to the moral and personal emotions, distorts and exaggerates the problems that we may face here.[[4]](#endnote-4) It is not obvious, for example, that the reactive attitudes should be taken to comprehend *all* the various emotions involved in personal relationships (e.g. love and friendship). If this is correct, then abandoning the reactive attitudes need not imply the loss of all personal relations and emotional attachments. Beyond this, it may be further argued that not all *reactive* attitudes would be discredited or undermined by skeptical arguments, since some may be retained or amended, consistent with skeptical commitments. In a similar vein, it may be argued that not only are there some moral emotions that are not reactive attitudes (e.g. moral admiration), there may also be reactive attitudes that do not involve *moral* expectations or standards. In all these ways it may be argued that Strawson’s effort to link skepticism with a pessimism based on the consequences of universally adopting the objective attitude is too simple and exaggerates the challenges that the skeptic faces.

On the other side of the coin, critics have argued that Strawson’s effort to “reconcile” the Optimist and the Pessimist (Strawson, 1962/2013: 63), and provide a suitably qualified defence of optimism on this subject, underestimates persisting sources of pessimism for his own position. In particular, there are issues of “history” and “luck”, which lie at the heart of the Pessimist outlook, that Strawson fails to properly address and engage with. This line of criticism has been strongly pressed by, among others, Gary Watson. Although Watson is not entirely unsympathetic to Strawson’s (“expressivist”) approach, he argues that Strawson neglects what may be described as “the historical dimension of the concept of responsibility” (Watson, 1987/2013: 106-07). We can appreciate this dimension of the problem, Watson suggests, if we examine some real and specific cases, where an individual’s upbringing and formative experiences have brutalized them and turned them into vicious criminals. When we consider the crimes of these individuals our reactive attitudes will, no doubt, be strongly aroused – especially in cases where none of the usual excusing or exempting considerations that Strawson has mentioned seem to apply. Nevertheless, when we turn our attention to the background conditions that shaped the criminal’s character and conduct, Watson suggests, we feel ambivalence and conflict, as we come to see the criminal not just as a victimizer but also as a *victim* (Watson, 1987/2013: 101). Our awareness of moral luck, in cases of this kind, can be *generalized* to all cases, showing that our reactive attitudes may be more sensitive to historical considerations – and skeptical pressures – than Strawson acknowledges.

***III. Dispelling the Bugbears and Bogeymen – Dennett’s Revisionary Optimism***

A very different approach to the problem of moral responsibility is presented in Daniel Dennett’s *Elbow Room*, a work that was published more than two decades after Strawson’s “Freedom and Resentment”. Dennett’s work has little to say about reactive attitudes and, instead, offers an account of the importance of rational self-control for our understanding of these matters. There are two closely related themes woven together Dennett’s discussion. First, he aims to provide an account of what it is with respect to free will that we value and care about and to show that we do possess those abilities and qualities that are worth wanting. Second, he aims to unmask and debunk the various skeptical arguments and analogies that attempt to show that our self-image as free and responsible agents is in some way illusory or mistaken. The first task supports his optimism and the second discredits the pessimism that lurks behind these skeptical arguments. His conclusions, he says, are “neither revolutionary nor pessimistic”. “They are”, he continues, “only moderately revisionary: the common wisdom about our place in the universe is roughly right. We do have free will. We can have free will and science too.” (Dennett, 1984: 19)

Dennett’s optimistic conclusion that free will is not an illusion rests on a clear statement of what it is that we care about or “deem worth wanting” (Dennett, 1984: 153).

What we want when we want free will is the power to decide our courses of action, and to decide them wisely, in light of our expectations and desires. *We want to be in control of ourselves, and not under the control of others.* We want to be agents, capable of initiating, and taking responsibility for, projects and deeds. All this is ours [Dennett has] tried to show, as a natural product of our biological endowment, extended and enhanced by our initiation into society.” (Dennett, 1984: 169 – my emphasis)

The real threats to human freedom, Dennett maintains, are not metaphysical (e.g. determinism) but threats we face due to control, manipulation and coercion by others. These real threats to human freedom do limit and erode our responsibility. We can, nevertheless, still distinguish between “responsible moral agents and beings with diminished or no responsibility” - a distinction that “plays a crucial role in the quality and meaning of our lives” (Dennett, 1984: 155). When it comes to holding others responsible, or taking responsibility for our own conduct and activities, we do not need to rely on some absolutist conception of (God-like) “Cosmic Responsibility” (Dennett, 1984: 158). On the contrary, we can make perfect sense of our associated attitudes and practices (e.g. punishment) in terms of the way they serve to discourage actions and activities that cause unwanted harms and injuries (Dennett, 1984: 158-64). In this respect Dennett’s system remains firmly committed to most of the pragmatic, forward-looking features of the Optimistic model that Strawson had earlier criticized.[[5]](#endnote-5)

What is distinctive about Dennett’s approach, and constitutes a substantial advance on earlier statements of Optimism or classical compatibilism, is his account of freedom in terms of our ability to recognize and respond to reasons (Dennett, 1984: Chp. 2). Once we appreciate the way in which *complexity* matters, Dennett suggests, we are better placed to unmask and expose the various “bugbears” and “bogeymen” that torment incompatibilists and their worries about the consequences of determinism. One issue that incompatibilists have worried about is that if we accept compatibilist conditions of responsibility we might still be manipulated by other agents. In reply Dennett argues that our “power of reflexive monitoring” enable us to detect “sneaky manipulators” or “evil tricksters” (Dennett, 1984: 37-38). Nothing about the thesis of determinism implies that we do not possess and exercise abilities of this kind. Along with the various “bogeymen” of manipulation cases (e.g. “the Nefarious Neurosurgeon”) there are a number of other “bugbears” that Dennett seeks to discredit. It has been argued, for example, that if determinism is true then we would controlled by nature or controlled by the past – and could not, therefore, be in control of ourselves. Dennett objects that claims of this kind misleadingly turn the environment and the past into “superhuman intelligences”, when obviously they are nothing of the kind. “Far from it being the case that we are completely under the control of our ancestors or our evolutionary past”, he says, “it is rather the case that that heritage has tended to set us up as self-controllers – lucky us” (Dennett, 1984: 72).

The opening chapter of Dennett’s *Elbow Room* presents a more general overview of the way in which incompatibilist pessimism has acquired such a strong psychological grip over us. Worries and anxieties about determinism, he suggests, are the product of “fearmongery” by philosophers, who have “conjured up a host of truly frightening bugbears” (Dennett, 1984: 4). These pessimistic “gloomleaders” have relied, Dennett claims, on various thought experiments that serve as “intuition pumps” designed to arouse some relevant negative response in us (Dennett, 1984: 12, 18). The fact is, however, that all these “intuition pumps” rely on misleading analogies and metaphors that do not deserve the respect and influence they have enjoyed (Dennett, 1984: 7). Dennett’s fundamental task in *Elbow Room* is to “exorcise” these bugbears and eradicate this groundless, “self-induced panic” (Dennett, 1984: 169-70).

Although Dennett is no doubt successful in discrediting some of the cruder and more misleading features of the various “bugbears” that he considers, critics may argue that he does not succeed in dispelling all the worries and anxieties that may grip us. Among the most important of the “bugbears” that Dennett examines are the issues of fate and luck. As we have noted, much of what Dennett says on these subjects repeats the familiar repertoire of classical compatibilism. Dennett takes the doctrine of fatalism to be the claim that all our deliberations and actions are causally ineffective and make no difference to the course of events (Dennett, 1984: 104-05). He goes on to point out that nothing about the thesis of determinism implies that this is the *universal* condition. The problem with this reply, as his critics see it, is that it is evasive and fails to engage with the deeper difficulties we encounter here. The relevant issue is not about the causal influence *of* the agent but rather the causal influences *on* the agent. The incompatibilist argues that there is another mode of fatalistic concern that arises from a backward-looking perspective. The question that concerns us, from this perspective, is whether or not the agent has *ultimate* control over her character and conduct. Do the qualities and features of her will for which we may praise or blame her *originate* with her? The essence of the incompatibilist view is that if determinism is true then *origination* fate is the universal condition and responsibility is impossible?[[6]](#endnote-6)

Dennett’s effort to deal with the bugbear of luck runs into a number of related difficulties. From an incompatibilist perspective concern about origination-fate is intimately connected with worries about “moral luck”. It is intuitively unjust, they argue, to hold agents responsible for aspects of their conduct and character they do not control. Agents subject to origination-fate are vulnerable to “moral luck” of just this kind. Dennett aims to deflate the luck objection by arguing that human agents are “not just lucky” but are, on the contrary, “skilled self-controllers” (Dennett, 1984: 94). Here again, Dennett’s critics are not convinced. While it is true that human agents are sensitive to reasons and able to guide their conduct in light of their reasons, given determinism it is also true that the specific capacities we have, the circumstances in which they are engaged, and the way they are actually exercised depend, in the final analysis, on external factors that the agent has no control over. In sum, contrary to what Dennett claims, powers of reason-responsiveness fail to provide immunity from the various forms of (moral) luck that incompatibilists are worried about.[[7]](#endnote-7)

Dennett, as we have noted, believes that his account of our (compatibilist friendly) powers of rational self-control are all we need in order to defuse worries about (evil) manipulators of various kinds (Dennett, 1984: 8,33-4,37-8,64-5). At the same time, Dennett also advances a theory of responsibility that emphasizes the way in which society can deploy the stance of “designers” and “engineers” with a view to controlling and directing an agent’s character and motivations (e.g. through the effective implementation of rewards and punishments) (Dennett, 1984: 156-65; see also 139-41, on robot design). These core arguments in Dennett’s work, it may be argued, pull in opposite directions and reveal deep tensions in his overall project in *Elbow Room*. Although Dennett claims to distance himself from the “chilly” world of B.F. Skinner’s *Walden Two* and the “vision of behaviourism” that lies behind it, his general position on this subject might well be described as neo-Skinnerian[[8]](#endnote-8)

***IV. Hard Incompatibilism / Optimistic Skepticism***

Both Strawson and Dennett, we have noted, aim to refute skepticism and offer compatibilist accounts of moral responsibility that are plainly more sophisticated than the simpler models proposed by classical compatibilism (Optimism). Whereas Strawson’s model rests on the role of reactive attitudes and excuses, Dennett emphasizes the importance of our powers of rational self-control.[[9]](#endnote-9) Whatever their differences, they are both agreed that, *if* we can discredit skepticism, then we can vindicate some reasonable degree of optimism about our general predicament. Failing this, skepticism implies a significant measure of pessimism. In recent years, orthodox views of this kind have been challenged by “Hard Incompatibilists” or “optimistic skeptics”.[[10]](#endnote-10) They argue that skepticism has none of the bleak or disturbing implications or consequences that are generally attributed to it by its critics (such as Strawson and Dennett). The most influential and substantive statement of this view is presented in the work of Derk Pereboom.[[11]](#endnote-11) Pereboom’s approach, as the (alternative) label “optimistic skepticism” suggests, also falls into two parts (Pereboom, 2001: xxiii). The first involves the set of arguments advanced in support of skepticism about moral responsibility. The second concerns his effort to justify an optimistic attitude on the basis of the skeptical claims advanced.

Our concerns in this paper rest primarily with Pereboom’s effort to justify optimism. There are, however, two features of his skepticism that should be noted. First, Pereboom’s skepticism is not only incompatibilist with respect to determinism, it is also incompatibilist with respect to *indeterminism*. Pereboom is “agnostic” about the truth of causal determinism but maintains that we lack moral responsibility either way.[[12]](#endnote-12) It is for this reason he refers to this general position as “hard incompatibilism”. The second point about Pereboom’s skepticism that should be noted is that it is *qualified or limited* in its scope. Pereboom makes clear that there are several different senses of “moral responsibility” and that he is specifically concerned with the one “that has been at issue in the historical debate” (Pereboom 2001: xx-xxiii; Pereboom 2007: 86-7; Pereboom 2013: 421; Pereboom 2014: 2-4, 6-7; Pereboom 2017: 121-22).[[13]](#endnote-13) This is, he suggests, the concept that involves “the notion of basic desert” (Pereboom 2007:122-23; Pereboom 2014: 2). The basic desert notion suggests that for an agent to deserve blame (or praise) she must “be able to exercise a certain kind of control in its production” (Pereboom 2001: 47; Pereboom 2007: 86-7). In particular, the agent must be the *source* of her decision in the right way, and it is this sort of agency that we lack (whether determinism is true or false). What motivates these concerns with the (ultimate) source of our decisions and conduct is that basic desert serves as the foundation of our retributive dispositions and practices.[[14]](#endnote-14) The skeptic can allow, Pereboom maintains, that there are *other* forms of free will and moral responsibility that agents may actually possess and that are obviously compatible with determinism (Pereboom 2001: xxi-xxii; Pereboom 2007: 86; Pereboom 2013: 421-22; Pereboom 2014: 2-3, 131, 134). Nevertheless, efforts to defend *compatibilism* in these terms are, Pereboom maintains, irrelevant to the traditional debate about free will and moral responsibility (Pereboom, 2001: xxi-xxiii; Pereboom, 2017: 121-22).[[15]](#endnote-15)

On the basis of arguments advanced to show that both compatibilist and libertarian views must be rejected, Pereboom goes on to show that the Hard Incompatibilist position that follows has few, if any, of the unattractive consequences that are generally attributed to it. In this respect his method bears some resemblance to Dennett’s, in that he aims to eradicate various “bugbears” that critics of skepticism have conjured up. If this can be done, we can vindicate optimism on *the basis of skepticism* (as opposed to its refutation). The title of Pereboom’s first book – *Living Without Free Will* – aptly describes this program. According to Pereboom the program that he proposes is not only feasible, it is an *attractive* ideal that we have good reason to implement and put into practice.

There are three major sources of pessimism that are frequently said to follow from the skepticism of the kind that Hard Incompatibilists defend. The first is that skepticism about moral responsibility would have catastrophic consequences for morality and society. Pereboom denies this and argues that not only can we legitimately continue to judge actions as morally right or wrong, we still have sufficient resources to respond to wrongdoers in an effective and appropriate manner. Not only can we use “moral admonition and encouragement” – stripped of any association with basic desert – we can also respond to “crime” in ways that treat the wrongdoer as needing treatment and cure rather than punishment (Pereboom, 2013: 435-7).[[16]](#endnote-16) When an agent (criminal) is judged dangerous then it may be necessary to “quarantine” her – but there is, Pereboom maintains, nothing morally objectionable about that.[[17]](#endnote-17)

Another potential source of pessimism, critics suggest, is that in a world stripped of basic desert there would be no “genuine achievement”, since no agent would *deserve* credit for their accomplishments. Pereboom denies that achievement is as closely tied to praiseworthiness as the objection supposes. Nor does Pereboom accept that our sense of self-worth rests entirely on features of ourselves that depend on deserved credit. In ordinary life we take pride in various features of ourselves that do not depend on our efforts in any way (e.g. intelligence, beauty, etc.). Moreover, qualities and features that do depend on our own will may still be valued even if it turns out that they are not *freely* willed. Nor is there any reason, according to Pereboom, to experience any sense of dismay even if our *moral* character is not something that we have produced or *deserve* praise or blame for. Given that these worries can be handled, we can be confident that our lives will continue to have worth and significance in the absence of deserved praise or blame.

A final set of pessimistic worries (bugbears) that Pereboom seeks to discredit concern the suggestion that if skepticism is correct it will result in the complete erosion and distortion of our personal relationships and emotional lives. This is a line of pessimistic concern that, as we noted, Strawson advanced in “Freedom and Resentment”. Contrary to what Strawson suggests, not all forms of reactive attitude will be discarded if we accept skepticism. There are some “analogues” that will be “sufficient to sustain relationships” (Pereboom, 2007:119). Moreover, many of those that do not survive, on skeptical principles, we would be better off without – such as (desert-based) anger and resentment (Pereboom, 2013: 443-4). Included in the range of emotions that are not endangered by skeptical principles or that may survive in an “analogue” form, Pereboom argues, are forgiveness, gratitude, mature love, regret and moral sadness. Taken together, all these emotions are sufficient to avoid anything resembling the bleak, dehumanized existence that troubles Strawson when he describes a world dominated exclusively by the objective attitude.

At this juncture there may be some questions we may as to whether there is any real *distance* between Hard Incompatibilism and Dennett’s compatibilism, beyond they way in which they frame or label their respective positions. Clearly both Pereboom and Dennett are skeptical about moral responsibility in the “basic desert sense” [herafter BDR] (Pereboom 2014: 2-3; cp. Pereboom 2001: xxi-xxii; Dennett, 1984: 166). Pereboom also accepts that there are *non*-BDR understandings of moral responsibility that are “compatible” with determinism – including the specific forms of moral responsibility that Dennett aims to describe and defend. They are also both agreed that jettisoning BDR and retaining surviving modes of responsibility allows for a generally *optimistic* attitude. On all these key issues, therefore, they converge. There is, perhaps, some real disagreement about the extent to which these shared conclusions “leave everything where it is” – especially as concerns the issue of punishment (although, even here, disagreement may be more apparent than real). Beyond this, however, the only remaining point of disagreement we are left with is that Pereboom, as we noted, objects to Dennett-style *compatibilism* on the ground that it presents other kinds moral responsibility as a *substitute* for BDR, which he takes to be an attempt to “recast the debate” (Pereboom 2001: xxii; and see also Harris 2011: 20-26). Pereboom objects, in particular, that presenting some *non*-BDR interpretation under the heading of “compatibilism” turns almost everyone into a “compatibilist” (Pereboom, 2014: 2; similar criticisms of compatibilists are advanced in Nadelhofer 2010). It may be pointed out, in reply, that this line of criticism turns every compatibilist who rejects BDR into a “skeptic”.[[18]](#endnote-18) For many compatibilists, however, including Dennett, the crucial point is that partial (qualified) skepticism about “absolutist” or BDR conceptions of moral responsibility does not imply *total* (unqualified) scepticism. Insisting that compatibilists should be labelled “skeptics” (*tout court*) obscures this point.[[19]](#endnote-19)

Setting this issue aside, the question still remains with us about how convincing it is to claim that skepticism about basic desert responsibility has none of the bleak and troubling implications or consequences that have been attributed to it? The basis for much of the *optimism* claimed by (partial) skeptics is that by pushing us to abandon the illusory self-image required by BDR we are encouraged to discard and repudiate cruel and destructive *retributive* emotions and practices. This is, no doubt, an attractive and humane aim of the Hard Incompatibilist position. It is not clear, however, that skepticism about BDR is needed to achieve this end. The underlying concern here is that the skeptic operates with a questionable picture of the relationship that holds between BDR and our retributive principles and practices. On the face of it, Hard Incompatibilists present BDR as committed to some form of *positive* retributivism, which suggests that blameworthy individuals who have committed a wrong *ought* to be punished or that it is good that they are made to suffer.[[20]](#endnote-20) There are, however, weaker and more restrained forms of retributivism, such as *negative* retribution, which requires only that we limit punishment to the guilty, and *permissive* retributivism, which allows that the guilty may be punished but does not demand this. If we take BDR to be committed to a *positive* retributive principle of some kind then it is, indeed, a severe and unattractive doctrine. It is, however, highly questionable that a notion of moral responsibility of this (narrow and extreme) kind represents *our* “ordinary” views – which are, arguably, much more conflicted, fragmented and unstable than this analysis suggests.[[21]](#endnote-21)

It may be argued that Hard Incompatibilism not only exaggerates the benefits of (partial) skepticism, it also tends to downplay or minimize its *costs*. In support of their view Hard Incopatibilists claim that there are available “analogues” and alternative to desert-based reactive attitudes and that these other forms of moral emotion will suffice to support and sustain a (post-BDR) morality. Critics may question how *effective* alternative or analogue versions of our reactive attitudes would be in supporting and sustaining (post-BDR) moral life? Consider, for example, the proposal that (moral) “sadness”, “disappointment” and “regret” can fill the void left by our desert-based reactive attitudes. Responses of this kind seem too tepid and tame to do the work of their originals – especially in more serious or grave contexts and situations.[[22]](#endnote-22) Even if we agree, with the Hard Incompatibilist, that criminal agents do not *deserve* blame or punishment – and many will, obviously, contest this – any surviving “analogues” that remain available will an unconvincing substitute. When our moral resources are reduced and impoverished in this way, our *commitment* to the (moral) values will be correspondingly weakened and eroded. Justified or not, this is not a comforting or stabilizing message for morality.[[23]](#endnote-23)

***V. Mixed Theories – Affirmation and Illusionism***

Hard Incompatibilism is firmly committed to optimism and maintains that the skeptical conclusion that it advances should be enthusiastically welcomed. There are, however, some closely related positions that argue that the skeptical claims being advanced demand more concessions to pessimism. Influential views of this kind have been presented and defended by Ted Honderich and Saul Smilansky. They are both skeptics about BDR but they do not accept that skepticism is as benign as the Hard Compatibilist would have us believe. The views that Honderich and Smilansky have advanced may be characterized as “mixed” theories, since they absorb both optimistic and pessimistic elements into their stance. Where they differ from each other is over the matter of the *degree* of pessimism that follows from skepticism about BDR and what practical measures need be taken in response to the skeptical assessment of our human predicament.

Honderich’s “Affirmation” theory is embedded in his general “theory of determinism”.[[24]](#endnote-24) Determinism, as Honderich interprets it, rests on the claim that our choosing and deciding is subject to ordinary (necessitating) causal regularities. This is a theory, he argues that is “strongly supported” and has certainly “not been shown to be false” (Honderich 2002: 90). The question he asks is what are the consequences of determinism, if it is indeed true? His answer to this question begins by pointing out that if determinism is true then the opposing view of “the philosophy of Free Will” is mistaken, since it presupposes indeterminism (Honderich 2002: 2,4,35,41-2,49,69,76). A key aim behind the Free Will view is that we need to understand responsibility in a certain way. More specifically, it requires that responsible agents have “a kind of personal power to *originate* choices and decisions and their actions” (Honderich 2002: 2 – emphasis in original). Absolute responsibility of this kind not only requires the falsity of determinism it also requires some sort of “ongoing entity” that possesses the “active power” required to produce or not produce a given action (Honderich 2002: 42,49). If we accept the truth of determinism, then we must reject the idea of Free Will, so understood.

According to Honderich, accepting the truth of determinism and abandoning the idea of Free Will certainly damages a range of things that we care deeply about. This includes, among other things, our personal feelings and attitudes to other people (e.g. resentment and gratitude), to the extent that these depend on a notion of origination and Free Will. These personal feelings and reactions are closely bound up with our *moral* feelings and responses (Honderich 2002: 97-100). If there are no active agents of the kind presupposed in our idea of Free Will, Honderich suggests, this will discredit our entire understanding of retributive justice and deserved punishment (Honderich 2002: 101-03, 133-41). It is relation to this issue that our sense of *dismay* in confronting the truth of determinism will be felt most acutely.

Honderich recognizes that compatibilists will resist these claims. The compatibilist will argue that the entire analysis, leading to dismay, rests on a mistaken understanding of what we mean by freedom. Freedom is not, they maintain, a matter of origination of some metaphysical kind but rather a matter of voluntariness (Honderich 2002: 98-9). As long as we keep in mind that free actions have a certain kind of causal history we can draw all the relevant distinctions that we need to make. Even if we accept the truth of determinism we can still distinguish, for example, between voluntary and involuntary conduct, and distinguish conduct due to violence, ignorance, or merely accidental, and so on. The compatibilist concludes from this that determinism changes nothing and leaves everything as it was (Honderich 2002: 97,99). The compatibilist, therefore, rejects dismay and instead adopts an attitude of “satisfied intransigence” (Honderich 2002: 97).

Both these opposing responses to the likely truth of determinism, Honderich maintains, are unsatisfactory. The root mistake, common to both, is that they assume that we have just one, single, ordinary idea of freedom and that any alternative account is either inadequate or irrelevant (or both). Contrary to this view, Honderich argues that we have *two* distinct ideas of freedom or free choice, one understood in terms of origination and the other in terms of voluntariness (Honderich 2002: 109,120). Any effort to arrive at a proper understanding in terms of just one of these two concepts will be one-sided and incomplete. The fact is that when we consider the consequences of determinism we experience a *dual response* of both dismay and intransigence, which results in “attitudinarian instability and discomfort” (Honderich 2002: 125-6).

The solution to this problematic situation is not just “intellectual” or conceptual but a matter of reconciling our conflicted *feelings and attitudes* (Honderich 2002: 114). This requires some concessions on both sides. Contrary to compatibilist accounts, we must accept that some of the features and aspects of human agency that we care about cannot be retained or supported, as they are not consistent with the truth of determinism. It is, for this reason, a mistake to suppose that accepting determinism “changes nothing” or “leaves things just as they are”. However, while we must not deny that there are things we value that we must give up, we should not underestimate what we still have. Determinism does not justify any sense of defeat or resignation and to this extent our sense of dismay can be contained, if not eliminated. The upshot of all this is an attitude of Affirmation, which is a “philosophy of life” that encourages us to “accommodate ourselves to the situation we find ourselves in - accommodate ourselves to just what we can really possess if determinism is true, accommodate ourselves to the part of our lives that does not rest on the illusion of Free Will” (Honderich 2002: 126).

One important respect in which Affirmation is in agreement with Hard Incompatibilism is that it is skeptical about BDR – which rests on the idea of origination and cannot be rendered consistent with the truth of determinism. Where Affirmation diverges from Hard Incompatibilism is that it remains firmly committed to a stance of dismay (or pessimism) – an attitude that Hard Incompatibilists aim to discredit or minimize. The crucial point here is that Affirmation refuses to conceal the source of dismay that determinism generates, as grounded in skepticism about the idea of Free Will or origination. From this perspective, the *undiluted* optimism of Hard Incompatibilism involves a form of denial about the real losses that we must face up to. At the same time, Affirmation also insists that we can accept our condition, and live with it, without relying on any illusions, as there are plenty of considerations of a more reassuring kind for us to call on. Nothing about the consequences of determinism implies that our lives will “go cold” or suggests that “everything is wrecked” (Honderich 2002: 125). Accepting the (likely) truth of determinism does not leave everything where it was, and certainly involves the loss of aspects or dimensions of our lives that we have reason to care about. There are, nevertheless, no tragic or unbearable consequences that flow from this predicament.

Another “mixed” theory that resembles Affirmation in several respects is presented by Saul Smilansky in the form of “Illusionism” (Smilansky 2000/2013). Smilansky is, like Honderich, a “dualist” about free will in the sense that he holds that, while it is true that we enjoy compatibilist-friendly forms of freedom (i.e. voluntariness and forms of control associated with it), we also lack control in the sense of origination or ultimacy. Smilansky does not deny that compatibilism has “partial validity” in that the forms of freedom it describes are indeed needed to support a “Community of Responsibility”. What he denies is that this sort of compatibilist freedom is *sufficient* for human ethical life (Smilansky 2000/2013: 384-5, 393). More specifically, agential resources of this compatibilist kind, he argues, cannot support and sustain a wide range of fundamental and essential ethical values and interests. Among the values and interests that Smilansky has particularly in mind are justice and respect (Smilansky 2000/2013: 386-88). The general difficulty we face is that while we need to believe in (libertarian) conceptions of freedom and responsibility, in order to maintain our confidence in and commitment to ethical life, this belief collapses under critical scrutiny. This theoretical problem, Smilansky maintains, presents us with a real and troubling *practical* problem.

Although Smilansky accepts much of Honderich’s analysis, he takes a much more pessimistic view about the consequences of determinism and the accompanying loss of our belief in Free Will (*qua* ultimacy and origination). As Smilansky sees it, Affirmation is not a viable or credible solution to the predicament that we actually face. For this reason, he proposes a much more radical solution – “illusion”. If we are to maintain a secure foundation for our core values and ethical practices this will require that the illusory belief in free will should be promoted and encouraged. Given that a belief of this general kind is “a condition of civilized morality and personal value” (Smilansky 2000/2013: 88) this is not something that we can afford to abandon or give up. Clearly, then, Smilansky’s (deep) pessimism about the implications of (openly) embracing skepticism about origination and ultimate agency leads him to advocate this more radical proposal. Smilansky is, of course, well aware of the both the considerable costs of this (e.g. with regard to truthfulness) and the challenging practical obstacles (e.g. maintaining patterns of deception by elites over the masses). He argues, nevertheless, that all this is necessary to avoid a catastrophic erosion of human ethical life and practice.

When we compare Hard Incompatibilism, Affirmation and Illusion with each other, they fall into a spectrum of attitudinal alternatives, and all of them are founded on skepticism about BDR. Hard Incompatibilism, as presented by Pereboom, Caruso, Waller, and others, embraces and welcomes skepticism as providing an opportunity for a radical transformation of our attitudes and practices, grounded on the skeptical arguments and reflections that have been advanced. Smilansky views these developments with a sense of dread and alarm and he encourages us to *conceal* this truth about our predicament, in the hope of averting the (ethical) damage that it is liable to bring about. Honderich takes a middle path between these two poles, arguing that although there certainly are bleak and troubling losses that we must accept if we discard our belief in Free Will, we have, nevertheless, plenty of other resources to call upon to support ethical and meaningful lives. One important consideration that separates Affirmation and Illusionism, on one side, from Hard Incompatibilism on the other, is that for the former pair of views the primary source of their *persisting pessimism* is rooted in skepticism about BDR. In contrast with this, Hard Incompatibilism presents our situation in a very different light, holding that our mistaken and illusory belief is mostly a source of human misery and irrationality, and one that we are better off without.[[25]](#endnote-25) Despite these differences all three parties are agreed that *some form* of compatibilist freedom and associated forms of moral responsibility survive skepticism about BDR. Considered in their own terms, these compatibilist forms of responsibility are understood to be in no way troubling or problematic – much less a source of pessimistic concern. According to these views, whatever sources of pessimism there may be, we do not encounter them here.

***VI. Free Will Pessimism***

We have noted that both libertarians and compatibilists generally accept the orthodox view that pairs skepticism and pessimism and, conversely, suggests that the defeat of skepticism serves to vindicate (some measure of) optimism. We also considered the Hard Incompatibilist’s unorthodox view that skepticism may be construed in optimistic terms. Mixed views, such as Affirmation and Illusionism, maintain the orthodox linkage between skepticism and pessimism, locating sources of pessimism in scepticism about BDR, while retaining differing measures of optimism based on compatibilist considerations. What has not been considered, so far, is the suggestion that arguments and considerations advanced that *defeat* skepticism about moral responsibility may, nevertheless, serve as a basis for pessimism. Given the standard matrix of options, this may be received as a puzzling proposal. It is, however, a view of this kind that is advanced and defended by the Free Will Pessimist.[[26]](#endnote-26)

The approach taken by Free Will Pessimists relies on distinguishing carefully between issues of scepticism and pessimism. (We have already noted, in relation to Hard Incompatibilism, that it is a mistake to conflate these issues.) Free will scepticism is the view that the lack of origination agency, along with vulnerability to moral luck, serves to discredit our view of ourselves as free and responsible agents (*tout court*). Free will pessimism rejects free will scepticism, since the basis of its pessimism rests with the assumption that we *are* free and responsible agents. All the major parties and positions in the (traditional) free will debate resist the suggestion that we are free and responsible agents who are, nevertheless, subject to significant modes of fate and luck. According to free will pessimism, all the major positions are, in different ways, modes of philosophical evasion that encourage an untruthful picture of the human predicament with respect to agency and ethical life.

At the heart of the free will debate, structuring the standard options that we are presented with, there is a fundamental shared assumption. What all parties are agreed about is that any *acceptable* account of conditions of responsible agency must be such that they do not allow that agents who satisfy these conditions may also to be subject to conditions of fate and luck. Strong resistance to this proposal is something that unites all the parties, whatever other disagreements there may be. Disagreement, across the compatibilist/incompatibilist boundary, is rooted in different understandings or interpretations of what the relevant (contested) notions do or do not involve. Compatibilists and libertarians propose very different accounts of the powers and abilities that are required to *exclude* the presence of fate and luck. Both sides contest the adequacy of their opponent’s proposals. The skeptic maintains that no satisfactory or credible “solution” to this problem can be found – but insists, nevertheless, that the relevant standard in place must still be *respected* and cannot be simply be discarded or dropped.

The main focus of critical attention for the free will pessimist is the (shared) assumption that any credible account of freedom and responsibility must *exclude* conditions of fate and luck. It is this fundamental assumption that the free will pessimist challenges, maintaining that there are available of conditions of free, responsible agency that fail this test but are, nevertheless, adequate for the purposes of sustaining and supporting ethical life and practice.[[27]](#endnote-27) Although accounts of this kind retain an appropriate *degree* of connection between conditions of control (e.g. as involving the capacity and exercise of rational self-control, under some interpretation), they reject the ambition to exclude or eliminate all surviving modes of fate and luck. Among the forms of fate and luck that would still be admissible, on this view, are limits relating to agent’s control over the circumstances in which she decides and acts, the background features that may shape and condition her motivation and character, and the particular way that the agent may exercise her deliberative and agential capacities. In conditions of this kind, the free will pessimist maintains, agents may be *both* (fully) free and responsible *and* still subject to significant limits of control over the moral trajectory of their lives. All the established parties involved in the (traditional) free will debate simply rule-out this possibility on the ground that it is a necessary condition of the adequacy of any proposed account of freedom and responsibility that conditions of this kind do not coexist.

The core objection to any general proposal of this kind is rooted in concerns about “fairness” and justice. From this perspective, it is neither fair nor fitting to hold agents who are limited in these ways responsible for their conduct. This criticism has force, the free will pessimist may reply, only if we aspire to some form of *absolute* fairness, that “runs all the way down”, and distinguishes morality from other forms of agential evaluation that do not rely on this (metaphysically ambitious) standard (Russell 2008). This absolutist standard places impossible metaphysical weight on the requirements of responsible agency, which is why it collapses into (complete, unqualified) skepticism. The free will pessimist rejects these aspirations along with the standards that they promote and encourage. The issue of fairness, on this alternative view, is *internal* to the fabric of ethical life and the norms and values that they track (e.g. as broadly described in Strawson 1962). These norms and values may well be modified, corrected and amended over time, on the basis of critical reflection. However, any demand for fairness beyond or above this – some mode of absolute fairness – is a product of an illusory self-image and an untruthful picture of our human ethical predicament.[[28]](#endnote-28)

With respect to the related objection that free will pessimism, as presented, fails to defeat or discredit skepticism about BDR or provide a “solution” to the free will problem as understood in these terms, the free will pessimist can readily accept this. It is not the aim of the free will pessimist to offer a solution to the “traditional” problem, much less to respect this particular standard of success. On the contrary, it is a matter of fundamental importance for the free will pessimist that these aims and aspirations, which structure the traditional problem, should be rejected and abandoned. They are faulty because they rule out the very *possibility* that we may encounter (fully) responsible agents who are, nevertheless, subject to conditions of fate and luck. The only skepticism that free will pessimist endorses is *partial* scepticism (i.e. as targeted on BDR and related concepts that aim to satisfy its demands). A skepticism of this *limited* nature can, however, still allow for the legitimacy of (robust) forms of moral responsibility that do not deny the possibility of responsibility in face of conditions of persisting fate and luck.[[29]](#endnote-29)

One way we can account for the *pessimism* involved in this stance is to consider free will pessimism in relation to what Bernard Williams calls “the *morality* system” (Williams 1985: Chp. 10). The morality system, as Williams describes it, places particular emphasis on the (peculiar) concept of obligation, along with the closely related concepts of blame and voluntariness. Moral responsibility, as the morality system understands it, is taken to be primarily a matter of rational agents voluntarily violating their obligations and, thereby, being liable to blame and retribution. Williams refers to these sub-features of *morality* as “the blame system” (Williams 1985: 217). Understood in these terms, the concept of BDR and of “accountability” may be viewed as specific modes or forms of “the blame system”. Building on these core concepts and features, the morality system demands that moral responsibility, understood in these terms, must somehow be capable of “transcending luck”, providing a “purity” that only genuine (rational) agency of some kind makes possible (Williams 1985: 217). Within this framework, the aspiration of libertarianism and its commitment to some form of absolute, ultimate origination is entirely intelligible. Although compatibilists plainly reject these metaphysical ambitions they generally remain committed to many of the features of “the blame system” and argue that it can be satisfied *within* compatibilist constraints.[[30]](#endnote-30) In contrast with views of this kind, free will pessimism rejects “the blame system” and the forms of moral responsibility that are essential to it. As the free will pessimist sees it, the aspirations of “the blame system” not only motivate the various philosophical conundrums that plague the free (traditional) will problem, they also distort and misrepresent our understanding of the human ethical predicament (which plainly lacks the sort of “purity” that the blame system aspires to and demands). We are, as Williams suggests, “better off” without “morality” and, in particular, better off without “the blame system”.[[31]](#endnote-31)

Once these distinctions are drawn it is apparent that in abandoning “the blame system” we abandon significant *optimistic* features that are integral to it. In particular, “morality” offers us a comforting story to tell about the human predicament in respect of moral agency - a story that runs deep in Western society. Morality puts up strong resistance to any disturbing or troubling view of human ethical life that suggests that the manner in which agents exercise and operate their moral capacities and abilities depends in large measure on features that are not controlled or governed by these same capacities and powers. As Thomas Nagel puts it, accepting an outlook of this kind would leave us “morally at the mercy of fate”. Nagel takes this to be a *reductio* of any outlook of kind, whereas the free will pessimist maintains that it is the only *truthful* description of our predicament available to us. Trying to avoid this truth about our circumstances and conditions is exactly what generates the (intractable) conundrums of the free will problem and the various forms of philosophical evasion associated with it.

We may conclude by noting, with Bernard Williams, that “philosophy, and in particular moral philosophy, is still deeply attached to giving good news” (Williams 1996/2006: 49). This tendency is particularly apparent and heavily pronounced in the debate concerning free will and moral responsibility. Not only do almost all the parties aim to deliver “good news” or an optimistic conclusion of some kind, the very assumptions and aspirations that structure and motivate the whole debate manifest this prejudice. In contrast with this general outlook, what free will pessimism claims is that with respect to these issues, what we have is not an intractable (philosophical) *problem* that is still looking for a solution but rather a troubling and unsettling *predicament* that must be recognized and acknowledged.[[32]](#endnote-32)

**Bibliography**

Caruso, Gregg. (2016). “Compatibilism and Retributivist Desert Moral Responsibility”. Erkenntnis: 82.4

-- (2017). “Origination, Moral Responsibility, and Life-Hopes”. In G. Caruso, ed. *Ted Honderich on Consciousness, Determinism and Humanity*. London: Palgrave.

-- (2018). "Skepticism About Moral Responsibility", The Stanford Encyclopedia of Philosophy (Spring 2018 Edition), Edward N. Zalta (ed.), URL = <https://plato.stanford.edu/archives/spr2018/entries/skepticism-moral-responsibility/>.

--- [Forthcoming]. “Free Will Skepticism and Its Implications: An Agument for Optimism”, In E. Shaw, D. Pereboom, and G. Caruso (eds.), Free Will Skepticism in Law and Society, Cambridge: Cambridge University Press.

Cuypers, Stefaan E. (2013). “Moral Shallowness, Metaphysical Megalomania, and Compatibilist Fatalism”. *Ethical Theory and Moral Practice*: 16.1, 173-188.

Dennett, Daniel. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting*. (Oxford: Clarendon Press).

--. (2003). *Freedom Evolves*. New York: Viking.

Fischer, John and Robert Kane, Derk Pereboom, and Manuel Vargas. (2007). *Four Views on Free Will*. Oxford: Blackwell.

Hart, H.L.A. (1959/1968). *Punishment and Responsibility: Essays in the Philosophy of Law.* Oxford: Clarendon Press.

Harris, Sam. (2012). *Free Will*. New York: Free Press.

Hartman, Robert J. (2017). *In Defence of Moral Luck*. London: Routledge.

Honderich, Ted. (1988). *A Theory of Determinism: The Mind, Neuroscience and Life-Hopes*. Oxford: Oxford University Press.

-- (2002). *How Free Are You?* 2nd ed. (Oxford: Oxford University Press).

Kant, Immanuel. (1781/1965). *Critique of Pure Reason*. Trans. Y N.K. Smith. New York: St. Martin’s Press.

Levy, Neil. (2011). *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. (Oxford: Oxford University Press).

Mackie, J.L. (1985). “Morality and the Retributive Emotions”. In *Persons and Values: Selected Papers Volume II*. Oxford: Clarendon Press.

McKenna, Michael (2012). *Conversation and Responsibility*. (New York: Oxford University Press).

Miles, James B. (2015). *The Free Will Delusion: How We Settled for the Illusion of Morality*. Kibworth: Matador.

Nagel, Thomas. (1976/2013). “Moral Luck”. Reprinted in Russell & Deery (2013), 31–42.

Nadelhoffer, Thomas. (2010). “The Threat of Shrinking Agency and free Will Disillusionism”. In W. Sinnott-Armstrong and L. Nadel, eds. *Conscious Will and Responsibility*. New York: Oxford University Press.

Nichols, Shaun. (2007). “After Incompatibilism: A Naturalistic Defence of the Reactive Attitudes.” *Philosophical Perspectives* 21: 405-28.

Pereboom, Derk. (2001)*. Living Without Free Will*. Cambridge, MA: Cambridge University Press.

-- . (2007). “Hard Incompatibilism”. In Fischer, Kane, et al., 2007: 85–125.

-- . (2013). “Optimistic Skepticism about Free Will”. In Russell and Deery, 2013: 421–449.

-- . (2014). *Free Will, Agency, and Meaning in Life*. Oxford: Oxford University Press.

--. (2017). “Skeptical Views about Free Will.” In K. Tempe, M. Griffith, and N. Levy, eds. *The Routledge Companion to Free Will*. New York & London: Routledge.

Russell, Paul. (2000/2013). “Compatibilist Fatalism: Finitude, Pessimism and the Limits of Free Will”. Reprinted in Russell & Deery (2013), 450–68. [Reprinted in Russell: 2017b.]

--. (2002). “Pessimists, Pollyannas and the New Compatibilism”, in Robert Kane, ed., *The Oxford Handbook of Free Will*. (New York: Oxford University Press), 229–56. [Reprinted in Russell: 2017b.]

--. (2008). “Free Will, Art and Morality”. *Journal of Ethics* 12: 307–25. [Reprinted in Russell: 2017b.]

-- (2017a). “Free Will Pessimism”. In *Oxford Studies in Agency and Responsibility, Volume 4*, edited by David Shoemaker. Oxford University Press: 93-120. [Reprinted in Russell: 2017b.]

-- (2017b). *The Limits of Free Will: Selected Essays*. (New York: Oxford University Press).

Russell, Paul, and Oisin Deery, eds. (2013). *The Philosophy of Free Will: Essential Readings from the Contemporary Debates*. (New York: Oxford University Press).

Schlick, Moritz. (1939). *Problems of Ethics*. Trans. By D. Rynin. New York: Dover.

Shabo, Seth (2012) “Where Love and Resentment Meet: Strawson’s Intrapersonal Defence of Compatibilism,” Philosophical Review. 121: 95-124.

Skinner, B.F. (1948). *Walden Two*. Toronto: Macmillan.

-- 1971. *Beyond Freedom and Dignity*. Harmonsworth: Penguin.

Smilansky, Saul. (2000). *Free Will and Illusion*. (Oxford: Oxford University Press).

-- (2001/2013). “Free Will: From Nature to Illusion.” *Proceedings of the Aristotelian Society* 101: 71-95. Reprinted in Russell and Deery (2013): 379-99.

## Sommers, Tamler (2007) “The Objective Attiude.” *The Philosophical Quarterly*. 57: 321–341.

Strawson, P.F. (1962/2013). “Freedom and Resentment”. Reprinted in Russell and Deery (2013): 63-83.

Strawson, Galen. (1994/2013). “The Impossibility of [Ultimate] Moral Responsibility”. Reprinted in Russell & Deery (2013): 363–78

-- 1998. "Luck Swallows Everything," Times Literary Supplement, June 26, 1998*.*

Wallace, R. Jay. (1994). *Responsibility and the Moral Sentiments*. (Cambridge, MA: Harvard University Press).

Waller, Bruce. (2011). Against Moral Responsibility. (Cambridge, MA: MIT Press).

Watson, Gary. (1986). “Review of Daniel Dennett*, Elbow Room: The Varieties of Free Will Worth Wanting”*. *Journal of Philosophy*: 517-22.

-- (1987/2013). “Responsibility and the Limits of Evil”. Reprinted in Russell & Deery (2013): 84–113.

Williams, Bernard. (1985). *Ethics and the Limits of Philosophy*. With a commentary by A. W. Moore and a foreword by Jonathan Lear. (London and New York: Routledge (Routledge Classics edition 2011)).

--. (1996/2006). “*The Women of Trachis*: Fictions, Pessimism, Ethics.” Reprinted in *The Sense of the Pas*t. Ed. With an Introduction by M. Burnyeat. (Princeton: Princeton University Press.

Abstract

We might describe the philosophical issue of human freedom and moral responsibility as an existential metaphysical problem. Problems of this kind are not just a matter of theoretical interest and curiosity, they address issues that we care about and that affect us. They are, more specifically, relevant to the significance and value that we attach to our lives and the way that we lead them. According to the orthodox view, there is a tidy connection between skepticism and pessimism. Skepticism threatens a wide range of interests and concerns that themselves rest on the foundation of our self-conception as responsible moral agents. From this perspective, whereas skepticism licences a degree of pessimism about our human predicament, the defeat of skepticism serves to vindicate optimism. In recent years this orthodox view of the relationship between skepticism and pessimism has been challenged. It has been argued, for example,  that skepticism may be defended in much more optimistic terms. While we have reason to accept skepticism, we have no reason to draw any bleak or depressing consequences from this. Another way of severing the orthodox connection between skepticism and pessimism is to reject skepticism but deny that this will serve to secure or salvage any unqualified form of optimism. This article reviews and contrast these various positions and approaches, beginning with an account of P.F. Strawson’s particularly influential statement of the relationship between the skeptical challenge and pessimism.

**Endnotes**

1. Kant, Immanuel. 1781/1965: II, ii, Chp.2 (esp. p. 635). [↑](#endnote-ref-1)
2. Strawson’s arguments have, of course, been criticized in a number of different respects which cannot all be reviewed here. One place to begin is McKenna & Russell, 2008. See also Russell, 1992; and Russell, 2010. [↑](#endnote-ref-2)
3. The Optimist who Strawson specifically cites is Nowell-Smith (1948) but there is good reason to suppose that his most obvious target, in this context, is Moritz Schlick (1939), who presented a particularly influential statement of the classical compatibilist position. It was Schlick’s view that free will is “a pseudo-problem” and the responsibility can be reduced to the question of when rewards and punishments are effective “educative measures” – a view that falls squarely into Strawson’s sights. Another important figure in the context is B.F. Skinner (1948), a leading figure in behavioural psychology. Skinner’s views are briefly discussed further below [note 8]. [↑](#endnote-ref-3)
4. See, e.g., Wallace, 1994: esp. Chp. 2; Pereboom, 2001: esp. 199-207; and Sommers, 2007. [↑](#endnote-ref-4)
5. In his review of *Elbow Room* Gary Watson suggests that “Dennett’s treatment of responsibility seems the least instructive part of the book” and suggests, in particular, that Dennett’s account could benefit if made more use of Strawson’s discussion of “personal relations” in “Freedom and Resentment” from (Watson 1986: 522). [↑](#endnote-ref-5)
6. For another perspective on this matter see Russell 2000/2013; and Russell 2002. [↑](#endnote-ref-6)
7. Nagel 1976/2013 provides the classical statement of these worries. For other similar lines of criticism see, for example, Miles, 2015: 58-74; and also Levy 2011; Waller 2011; Harris 2012. [↑](#endnote-ref-7)
8. In *Walden Two* (1948) Skinner employs the apparatus of his “behavioural science” to describe a (dis)utopian world in which the benefits of the advances in the science of human behaviour could be taken advantage of with a view to creating a superior ethical agent and citizen. Dennett is careful to reject Skinner’s language of “man being controlled by his environment” (Skinner 1971: 19-20; Dennett 1984: 57-9), and is certainly aware of the potential misuse of behavioural control (Dennett 2003: Chps. 9 and 10). It may be argued, nevertheless, that like Skinner he *welcomes* the prospect of the state or authorities “engineering” society and controlling others by means of a “technology of behavior”. While we may be skeptical about the prospects of “behaviourism”, as Skinner conceives it, worries of this more general kind are not just science-fiction but constitute a *real threat*. (Consider, for example, some of the policies and practices, employing new techniques and scientific methods relating to mind-control with a view to create “better citizens” are now being implemented in China, among other places.) Belief in “autonomous man” may well be illusory but when it is discarded, we must confront some *uncomfortable* problems. [↑](#endnote-ref-8)
9. The respective strengths of each theory could be combined into a unified account. See, e.g., Wallace, 1994. [↑](#endnote-ref-9)
10. These labels are used by Pereboom and others in this school have followed him. [↑](#endnote-ref-10)
11. Other important figures who also belong to this school include Waller, 2011; Harris, 2012; Miles, 2015; Nadelhoffer 2010, and Caruso, 2016, Caruso, forthcoming. For useful background see Caruso, 2018. [↑](#endnote-ref-11)
12. Unlike some skeptics (e.g. G. Strawson, 1994), Pereboom does not claim that moral responsibility is “impossible”. What he claims is that although being “undetermined agent causes” may be a coherent possibility, it is not credible given our best theories”. (Pereboom 2013: 422).] [↑](#endnote-ref-12)
13. Given that the form of skepticism that Hard Incompatibilist defend is “targeted” in this way, we may describe it as a *partial* skepticism, to distinguish it from a more radical, *total* skepticism. However, in light of this distinction, it may be argued that there is a latent instability in the Hard Incompatibilist position about whether the skepticism that they defend is partial or total. To the extent that Hard Incompatibilists insist that the traditional, historical debate is (entirely) about basic desert responsibility, and that compatibilists are, therefore, really “skeptics” (e.g. Pereboom 2001: xxi-xxii; Pereboom 2014: 2-3; Caruso 2017: 200) they tend to blur this important distinction. For compatibilists it is crucial that this distinction should not be blurred in this manner, since partial skeptics are not committed to total scepticism. [↑](#endnote-ref-13)
14. Waller places particular emphasis on this point (Waller 2011: 2-7). According to Waller “the core concept of moral responsibility is what justifies blame and praise, punishment and reward; moral responsibility is the basic condition of giving and claiming both positive and negative deserts.” See also Miles, 2015: 42-44; Harris, 2012: 55-60; and, especially, Caruso, 2016). Caruso’s paper is devoted to providing a more detailed picture of “what the basic desert sense of moral responsibility amounts to”. He argues that it is tied very tightly with retributivist aims and values and suggests that it should be called “retributivist desert moral responsibility”. [↑](#endnote-ref-14)
15. Sam Harris devotes a whole chapter of his (brief) book on free will to this general concern, arguing that “compatibilists” are guilty of “changing the subject” (Harris 2011: 15-26. [↑](#endnote-ref-15)
16. It has been argued by some that it is not possible to separate wrongness from desert and retribution. For example, Mackie claims that “if we analyse the concept of a morally wrong action it will be seen to have much wider significance. In the central cases, the wrongness of an action is thought of as being made up of three elements, its being harmful, its being forbidden, and its calling for a hostile response. No one of these elements, if the others are somehow denied is sufficient to make an action wrong in the full sense” (Mackie, 1985:213). [↑](#endnote-ref-16)
17. A related worry here is about what *methods* may be used to “cure” or “help” the offender and protect society? What *limits* are there on such methods once we abandon the idea that we need to distinguish between (truly) responsible agents and those who are simply “ill” or “sick”? The scope for (arbitrary) interference and mind-control of various kinds would appear to be much more open-ended once these skeptical principles and assumptions are adopted or put into practice. This may well also be a problem for compatibilist theories, such as Dennett’s, but that does not show that the *skeptic* can deal with these concerns or has found a way around them. [↑](#endnote-ref-17)
18. From the perspective of Pereboom, Harris and other like-minded skeptics, those who identify themselves as “compatibilists” fall into two classes. There are those who accept BDR and (mistakenly) believe that it can be secured within compatibilist constraints. There are also those who reject BDR and are, therefore, really *skeptics* – since presenting themselves as compatibilists involves “changing the subject”. Suffice it to note, beyond this, that which camp compatibilists are assigned to will depend on how exactly we interpret the requirements of BDR (about which there is little consensus). [↑](#endnote-ref-18)
19. Accepting this, of course, is entirely consistent with rejecting any particular compatibilist account (including Dennett’s), which may well be judged inadequate or incomplete. Plainly compatibilists disagree among themselves as to what kinds of alternative accounts are adequate to the (complex) facts and standards that we need to appeal to [↑](#endnote-ref-19)
20. These distinctions are concerning retributive principles are drawn and explained in Mackie 1985: 207. [↑](#endnote-ref-20)
21. For a rather different perspective on retributivism and the various distinctions that are required to make sense of it, see Hart, 1959/1968: esp. Chp.1. The deeper problem here is that Pereboom’s account of “basic desert” is described largely in negative terms, as being not derived from consequentialist or contractualist considerations (on this see McKenna, 2012: 121). For this reason it remains unclear exactly how or what way BDR is related to retributivist principles. McKenna provides an illuminating examination of a number of the issues arising in relation to this general problem (McKenna, 2012: 4-5, 114-21, 150-70). [↑](#endnote-ref-21)
22. For criticism along these lines see Nichols 2007; but see Pereboom’s replies in Pereboom 2014: 146-49, 180-81. [↑](#endnote-ref-22)
23. It may be further argued that skepticism, if put into practice, is liable to result in collateral damage, for both our moral motivation and our appreciation of our various personal and moral relationships. The expectation that our moral and personal lives will carry on, largely undamaged, with few significant costs to bear for such a radical switch in our attitudes and self-understanding, seems too optimistic. (On this see Shabo, 2012; and Pereboom’s response in Pereboom, 2014: 183-86.) [↑](#endnote-ref-23)
24. Honderich’s theory of determinism was first presented in Honderich 1988 and then, in a revised and compressed form, in Honderich 2002. Although the later work “follows the same path” as the earlier work, is intended to be more than a mere *précis* of it. Readers should consult Honderich’s first and larger work for more detailed arguments and discussions relating to his views. [↑](#endnote-ref-24)
25. Thomas Nadelhoffer has described the Hard Incompatibilists as “the revolutionaries” in this debate and suggests that this outlook “might helpfully be called *free will* *disillusionism*” (as an obvious point of contrast with Smilansky’s “illusionism”). Nadelhoffer is unconvinced by claims relating to “the alleged benefits of believing in libertarian free will and moral desert” and declares his allegiance to the revolutionary cause (Nadelhoffer 2010). [↑](#endnote-ref-25)
26. Russell 2017a; see also Russell 2000; Russell, 2002; Russell, 2008; and Cuypers, 2013. [↑](#endnote-ref-26)
27. As I explain further below, this does not mean that “nothing changes” or that “the common wisdom about our place in the universe is roughly right” (Dennett 1985: 19). [↑](#endnote-ref-27)
28. For a valuable discussion and examination of the various modes of moral luck that may concern us in relation to these matters see Hartman 2017. [↑](#endnote-ref-28)
29. As we have already noted, there may well be disagreement and debate about the adequacy of various proposed alternative accounts. There is, however, no basis for dismissing *all* such alternatives accounts as inadequate, *unless* we impose the (contested) standard of BDR itself. [↑](#endnote-ref-29)
30. This aim is explicitly stated in Wallace 1994: 39-40, 64-6. Many would interpret Strawson as sharing these aims and aspirations (Strawson 1962). Even Dennett, who takes a self-consciously pragmatic turn on this subject, continues to cater to central features of “the blame system” – including the aspiration to “transcend luck”. [↑](#endnote-ref-30)
31. Williams 1985: 193. For reasons already touched on, it should be clear that abandoning “the blame system” and the (peculiar) forms of moral responsibility it seeks to impose on us, does not commit us to *total* or *complete* skepticism about moral responsibility. On the contrary, while adherents to the morality system tend to present the situation this way, it is generally recognized - even by “morality’s” adherents - that the narrow conception of moral responsibility, as constructed around this system, is one that is both “local” (i.e. modern and Western) and widely contested – even within our own modern, western ethical community. (See, e.g., Wallace 1994: 39-40, 64-5.) [↑](#endnote-ref-31)
32. I am grateful to Derk Pereboom for his comments and suggestions on an earlier draft of this paper. [↑](#endnote-ref-32)