



The Limits of Free Will

Paul Russell

<https://doi.org/10.1093/oso/9780190627607.001.0001>

Published: 2017

Online ISBN: 9780190627638

Print ISBN: 9780190627607

FRONT MATTER

Introduction

<https://doi.org/10.1093/oso/9780190627607.002.0009> Pages xiii–xxiv

Published: October 2017

Subject: Metaphysics, Philosophy of Mind

Collection: Oxford Scholarship Online

I

This collection is composed of a selection of chapters that represents work I have done on the topics of free will and moral responsibility. Most of the chapters have appeared over the past decade or so but I have also included a few older chapters that continue to be relevant to the current debate and discussion. The issues and problems addressed in these chapters are, of course, deeply rooted in the history of the subject and concern matters that are of perennial interest and importance in philosophy. The various issues addressed are not only intimately related to each other, but also of immediate relevance to neighboring fields, including subjects such as law and criminology, theology, moral psychology and, more recently, neuroscience. During the period that these chapters were written and first published this area of research has become increasingly active and vibrant. It has expanded and evolved well beyond the narrow and restrictive confines established by the methods and techniques of “analytic” philosophy as it was understood and practiced in the middle decades of the twentieth century. Although this collection does not reflect every aspect of the field, and some important problems and issues are not covered, most of the significant developments and changes that have taken place are well represented.

Among the various ways in which the contemporary debate has advanced and made progress there are four that are especially significant for the purpose of the chapters in this collection. First, the specific forms and varieties of compatibilist and libertarian positions have evolved and developed enormously over the past three or four decades.¹ The positions taken are not only more subtle and complex, they deal more effectively and convincingly with the familiar problems and objections. Even though these accounts have not settled or put an end to the debate, they serve as clear evidence of the degree to which our understanding and evaluation of these matters has advanced. Second, consistent with developments elsewhere in philosophy, there is a greater concern to offer accounts or theories that are empirically better grounded and more informed. This trend is apparent in libertarian theories but it is especially pronounced in contemporary compatibilist thinking and theorizing. This “naturalistic turn” in recent work has, among other things, taken the form of insisting on the particular relevance of moral psychology for our understanding of the normative framework within which these issues and problems arise. No serious theory or account of free will and moral responsibility can simply ignore the way in which justificatory issues are themselves embedded in and structured by the psychological attitudes and dispositions that are at work in this sphere.²

p. xiv

p. xv

The other two important developments are closely connected to each other, as well as being related to the two developments just described. One of these is the increasing seriousness with which the skeptical challenge to the whole edifice of free will and moral responsibility is now taken (e.g., Pereboom 2001; Waller 2011; Levy 2011; Harris 2012; Miles 2015). Only a generation or so ago it would be rare to find any serious theorizing about free will and moral responsibility from a skeptical point of view—the real debate was between the libertarian and compatibilist positions. This is no longer the case and one question that must, therefore, be taken more seriously is what is the significance of skepticism and what is involved in abandoning our self-image as free and responsible agents? Can we, for example, really *live* our skepticism in practice—is this something human beings are even capable of (Strawson, 1962)? The fourth development, related to the issue of skepticism, concerns our metaphysical attitudes of optimism and pessimism. It has been widely held that skepticism would commit us to a bleak and troubling understanding of the human predicament with respect to these matters and that, conversely, the defeat of skepticism serves to vindicate a more optimistic view. Several recent defenses ↵ of the skeptical view have, however, denied this linkage or association between skepticism and pessimism and have argued, instead, that the skeptic can sustain a coherent and plausible optimistic picture of human life, without the illusions and confusions involved in regarding ourselves as free and responsible agents (see, e.g., Pereboom 2001; Honderich 2002; Sommers 2007; Waller 2011)

The four developments outlined above provide the broad framework within which the chapters included in this collection have been written. Each chapter draws upon one or more of these developments and, taken together, they reflect these important changes in the free will debate over the past few decades. There are, of course, some significant and substantial developments in the contemporary debate that are not covered or discussed in these studies. For example, over the past two or three decades—as with philosophy more generally—considerable effort has been made to include and accommodate the findings of science into this debate. The sort of “armchair” approach that was encouraged by the methodology of analytic philosophy is now generally resisted and repudiated. A number of prominent philosophers and scientists have attempted to employ and apply the methodologies and data of science as a way of resolving the free will problem. Some interesting and stimulating research has certainly been generated along these lines—especially as it concerns the experiments and discoveries of neuroscience and social psychology. At the same time, this process has also encouraged some extravagant and excessive claims. This includes the skeptical claim that we have (decisive) “empirical evidence” to prove that free will is an illusion and there is no basis for moral responsibility, and so on. Just as some scientists and philosophers have embraced and endorsed these extravagant claims, others have played an important and valuable role in challenging and discrediting them. My own work, however, has not engaged in any detail with these particular debates, interesting and influential as they may be. One reason for this is that I am less enthusiastic—and perhaps more skeptical—than most of my colleagues about the extent to which these controversies, in the final analysis, serve to clarify or elucidate the core issues. I believe that we need to look elsewhere if we are to make significant and substantial progress on these issues and problems.

p. xvi

One way in which my own methodological approach differs from the dominant contemporary paradigms is that my approach to these issues has been deeply embedded in and combined with historically oriented studies. In contrast with this, contemporary investigations not only tend to place heavy emphasis on the techniques and findings of science, but ↵ also generally neglect or even dismiss the value of historical studies and sensibilities in relation to this topic (this being the reverse side of an attitude that relies narrowly on the model and methods of science).³ With this in mind, it is worth pointing out that the chapters that are included in this collection not only are intimately connected with these historically oriented studies of the free will problem, but also in a number of cases they have arisen directly from them. This relationship is especially obvious and pronounced with regard to several studies that I have presented concerning Hume’s philosophy and the way in which it is of (multiple) relevance to the contemporary free will debate (Russell 1995; Russell 2015). None of this is flagged in detail in the chapters that follow but it

should be evident to those who are familiar with the philosophy of Hume, along with the other historical figures I draw from.

With regard to the set of chapters included in this collection some are mostly critical in character, presenting critiques and commentary on major works or contributions in the contemporary scene. Others are primarily constructive, aiming to develop and articulate an alternative compatibilist theory—a theory, as I explain further below, that is deeply rooted in Strawson’s naturalistic program but diverges from it (and other important and more recent variants of it) in significant respects. There is, nevertheless, no simple division between the critical and constructive tasks, as these two aspects of my work are fused together and serve to unite the collection into a coherent whole.

Finally, each chapter in this collection is self-standing and can be read in isolation from the others. There is, nevertheless, a core set of themes and issues that unite and link them all together. This collection is arranged and organized in a format that enables the reader to appreciate and recognize these links and the core themes that unite them. This is a key rationale or justification for the whole project. The collection, considered as a whole is, I believe, much more than simply the sum of its parts. Presented in this format, the collection reveals the deep and significant structural relationships that hold between the discussions in these chapters and shows how they are relevant and connected to each other—something that is otherwise obscured, even for the careful and interested reader.

II

p. xvii

It may be helpful if I provide a few further details about the structure and content of this collection. This volume is divided into four parts, with twelve chapters in total. Each part is arranged around a core theme and most of the themes and chapters overlap and are interrelated in terms of their content and concerns. (These themes are related to but distinct from the four developments described in the section above.) Among the key topics taken up are the relevance of the metaphysics of causation for free will; the nature and credibility of the (Strawsonian) naturalistic program, which draws on the role of moral sentiment considered as a solution to the free will problem; skepticism about practical reason; the relevance of free will for art and morality; the adequacy of reason-responsive theories of moral agency; the manipulation argument; and, finally, pessimism about the limits of agency.

The following is a brief summary and review of each of these four parts and of the particular chapters that have been assigned to them.

I. Free Will and Causal Relations

There is an obvious sense in which the metaphysics of causation is central to the problem of free will and, according to some, it can serve as the relevant basis for its solution. The basic idea common to these approaches is that the difficulties we encounter in this area have been generated by faulty assumptions about the nature of causation and how it relates to human action and conduct. The solution, therefore, rests with identifying and removing the source of the relevant metaphysical confusion. It is significant that both libertarians and compatibilists have sought solutions along these lines—despite the very different conclusions that they aim to draw from it. In the two chapters that are included in this part of the collection I take up an approach from each side of this debate—both of which have been extremely influential and continue to command support. I argue that neither strategy is successful. The failure of approaches of this kind suggests that a satisfactory solution likely has to be found elsewhere.

Lying at the heart of these approaches to the free will problem through the metaphysics of causation is the question about the nature of causal *relations*. Two questions that are especially important are: (1) do causes always necessitate their effects? and (2) do causal relations involve metaphysical powers or forces of some kind or are they to be analyzed simply in terms of regularities or constant conjunctions of events? Over the past few decades a number of libertarians have advanced an “event-causal” theory that grounds libertarian metaphysics in explanatory but non-necessitating causal relations. The most prominent representative of this strategy is Robert Kane, who has presented and defended a particularly subtle and detailed theory along these general lines (Kane 1996). However, a similar view was also advanced and defended, a few years earlier, by Richard Sorabji (Sorabji 1980). In developing his own account, Sorabji drew on his interpretation of Aristotle. In the first chapter I offer a critique of Sorabji’s event-causal theory and argue that it runs into difficulties and objections located on both sides of the dilemma of determinism. I have included a brief “addendum” to explain the relevance of my critique to the contemporary debate (i.e., particularly as it concerns the debate around Kane’s model). The second chapter offers a critique of the classical compatibilist strategy, the central features of which originated with Hume but were further developed and defended by leading figures of nineteenth- and twentieth-century empiricism, such as Mill, Russell, Schlick and Ayer. The aim of this strategy was to dissolve the (pseudo-) problem of free will by way of diagnosing incompatibilist concerns as rooted in confusions about the nature of causation. I argue that these efforts to ground the compatibilist position in the regularity theory of causation not only manifest confusion about the original insights relating to the caused/compelled distinction, which is central to the compatibilist approach, but also threaten to generate some awkward problems for compatibilism by eroding the metaphysical bonds or ties between agent and action.⁴

II. Responsibility, Skepticism, and Moral Sentiment

Arguably the most important and influential contribution to the free will debate in the last half of the twentieth century has been P. F. Strawson’s “Freedom and Resentment” (Strawson, 1962). The second group of chapters pursues themes and issues arising out of Strawson’s contribution and the substantial discussion and debate it has generated. Strawson’s approach involves turning away from the more familiar and dominant debates of the twentieth century concerning alternative possibilities and the interpretation of the concept of freedom, to a close examination and description of the role that reactive attitudes or moral sentiments play in circumstances where we hold agents responsible for their actions. (These are concerns that follow similar lines of argument that can be found in Hume’s philosophy, which has been the focus of much of my attention in my more historical studies.) In “Strawson’s Way of Naturalizing Responsibility,” I am concerned with a key aspect of Strawson’s strategy; namely, his claim that no reasoning of any sort could lead us to abandon or suspend our commitment to the “reactive attitudes.” This is a claim that aims to discredit all radical, skeptical conclusions about moral responsibility based on concerns about the implications of determinism. I argue that Strawson fails to distinguish two very different forms or modes of naturalism and that he is constrained by the nature of his own objectives (i.e., the refutation of skepticism) to embrace the stronger, less plausible, form of naturalism. The critique provided suggests that there are significant gaps in Strawson’s effort to reconstruct compatibilism along these lines and that while his (neo-Humean) strategy has significant merits, it is not acceptable or convincing as it stands.

The chapter that follows, “Responsibility and the Condition of Moral Sense,” presents a thesis about necessary conditions of responsible agency that arise at the interface between (compatibilist) reason-responsive theories and Strawsonian naturalistic approaches. Contemporary compatibilists have suggested that Strawson’s theory has a significant gap, that it lacks an adequate theory of moral capacity (a point I also argue for in “Strawson’s Way”). A number of these critics have tried to plug this gap with an account of rational self-control or reason-responsiveness that does not involve any reference to moral sentiments and

our ability to *hold* agents responsible. The thesis argued for in this chapter is that the responsible agent (i.e., one who is capable of *being* responsible) must also be one who is capable of holding herself responsible. Where moral sense is lacking, I maintain, rational self-control is seriously impaired or compromised. The third chapter in this part of the collection is a critical summary and account of Strawson's approach to moral responsibility and free will, along with a discussion of several of the key responses it has generated. After reviewing Strawson's core arguments and identifying several significant weaknesses in it, I turn to R. Jay Wallace's effort to recast Strawson's program and amend and modify it in ways that avoid various objections that have been directed against it (Wallace 1994). Wallace's theory involves two strands, a Strawsonian account of holding people responsible and a Kantian account of moral agency. I argue that both these strands run into major difficulties. My discussion then turns to an analysis of Angela Smith's distinction between being and holding responsible, which she uses to criticize the whole Strawsonian program (Smith 2007). Contrary to Smith I argue, building on the discussion in the previous chapter, that there is a more complex and intimate relationship between these two aspects than her critique allows for or can accommodate. My discussion concludes with reflections on the relevance of an agent's history for our reactive attitudes and moral sentiments and whether or not such considerations license skeptical conclusions about the possibility of moral responsibility (e.g. as per Derk Pereboom's "hard incompatibilism"). In this context I also provide a sketch of "critical compatibilism" and "free will pessimism" (which are discussed at greater length in various chapters in Part IV).

The last chapter in Part II returns to a critical discussion of Wallace's effort to provide a "narrow construal" of moral responsibility in terms of a modified Strawsonian approach. My particular concern in this chapter involves an elaboration and expansion of my earlier critiques. While I endorse and share many of Wallace's objections to Strawson's way of naturalizing responsibility, I reject his effort to reconstruct and compress our understanding of moral responsibility into the restrictive framework of what Bernard Williams has described as "the morality system"—an understanding that places heavy and exclusive emphasis on the notions of obligation, voluntariness and blame in accounting for moral responsibility. In opposition to the narrow construal, I suggest we should embrace a broader conception that can accommodate more varied modes of moral sentiment and the diverse forms of moral responsibility that go with them. One of the significant benefits of taking this route is that it blunts or deflates the (global) skeptical challenge to responsibility, along similar lines to Strawson's original program.

III. Practical Reason, Art, and Manipulation

p. xxi

The chapters grouped together in Part III are more loosely related than the other parts of this volume but they are, nevertheless, thematically connected with the volume as a whole. The issue of practical reason is closely connected with problems of freedom and responsibility, since different views about the nature of moral freedom and moral capacity presuppose very different accounts of the powers of practical reason and its manner of operation and influence. The first chapter in this part of the collection takes up a crucial debate between Christine Korsgaard and Bernard Williams on this subject. Korsgaard defends a broadly Kantian view of practical reason against William's modified Humean view. It is Korsgaard's particular concern to argue that if reason can itself identify substantive ends for our actions, independent of our existing desires, then there is no genuine or distinct motivational problem about how reasons can move (rational) agents to action. In this chapter I argue that Korsgaard's argument fails and the motivational problem cannot be resolved along the lines that she proposes. In "Free Will, Art, and Morality" I consider the relevance of the free will problem for our evaluation of artistic achievement and merit. I argue that with respect to both issues, creativity and merit, incompatibilist worries about the implications of determinism are groundless and misplaced. On this basis I consider the implications of these conclusions in light of the significant analogies that hold between art and morality. I argue that whether incompatibilists accept or reject the analogy they face a series of intractable problems and dilemmas that tell against the incompatibilist position. The implications of all this for compatibilism, however, are in no way "comfortable" or confirming of "optimism." On the contrary, the relevance of the art and morality analogy highlights the extent to which moral evaluation is inescapably permeated by background conditions of luck. In this respect, there is an important sense in which we may say morality is unfair. The last chapter in Part III takes up a well-known objection to the compatibilist position, which is "the manipulation argument" and related arguments based on covert control. In this chapter I reject soft compatibilist responses to cases of this kind, which rely on considerations of "history" to exclude manipulated or covertly controlled agents from responsible agency. Instead I defend a modified form of hard compatibilism, one that grants there is something problematic about cases of this kind but rejects the claim that agents in these conditions are not responsible. The issue is not, I suggest, that these agents are not responsible but that their manipulators or covert controllers are not entitled to hold them responsible. In this way, selective hard compatibilism maintains that what is compromised in these circumstances is not the moral responsibility of the agent (where robust compatibilist conditions of a relevant kind are satisfied) but the participant stance or moral standing of their manipulators. It is these distinct considerations that account for the (limited) intuitive force of incompatibilist counterexamples of this general nature.

The previous parts of this collection include several contributions that suggest some proposals for how a broadly Strawsonian, compatibilist approach to free will and moral responsibility should (or should not) be amended and modified, with a view to meeting various objections that may be leveled against it. The final group of chapters in this collection, building on this, aims to articulate a distinctive account of compatibilism. Although it rejects any form of unqualified or radical skepticism, critical compatibilism insists that a plausible compatibilism has significant and substantive implications about the limits of agency and that this licenses a metaphysical attitude of (modest) pessimism on this topic. The first chapter, “Compatibilist Fatalism,” argues that compatibilists require a richer conception of fatalistic concern, one that recognizes the legitimacy of (pessimistic) concerns about the origination of character and conduct. On this basis I argue that compatibilists must allow that determinism has fatalistic implications of a significant and relevant kind, even if they are not responsibility undermining. This mode of fatalistic concern, I maintain, licenses a distinct form of pessimistic concern grounded in concerns about the limitations or finitude and contingency of human agency. No credible form of compatibilism can hope to evade this form of pessimism and, thus, all forms of compatibilism that aspire to metaphysical optimism in this respect are guilty of evasion and superficiality. In the chapter that follows I pursue this general line of thought with particular reference to the “new compatibilism” and theories of reason-responsiveness associated with it (Dennett 1984 is a particular target of my criticism). I argue that although compatibilism, in its various forms, may defeat immediate skeptical threats relating to the powers of agents to guide and control their conduct in light of reason and deliberation, we eventually reach a point where the way in which these powers and abilities are acquired and exercised falls outside the agent’s (ultimate) control. These reflections and observations, I argue, serve to justify “pessimism at the horizon.” This mode of pessimism is not, however, as explained before, rooted in skepticism about freedom and responsibility but rather in our understanding of the way in which the abilities and powers involved in the operation and exercise of rational self-control are themselves limited and reveal us to be agents who are subject to finitude and contingency, a reflection that is at least disconcerting, although not one that justifies any form of deep despair.

p. xxiii The last chapter in this volume weaves together the core arguments relating to critical compatibilism (which was also briefly mentioned above in the summary of Part II). This chapter draws a basic distinction between “free will skepticism” and “free will pessimism.” While any acceptable form of compatibilism cannot be skeptical it has, nevertheless, pessimistic implications relating to fate and luck. The source of strong resistance to this conclusion is rooted, I suggest, in “the morality system” and its aspiration to metaphysical optimism. Any plausible form of compatibilism must embrace free will pessimism and take the form of critical compatibilism (i.e., reject the optimistic aspirations of complacent compatibilism). Incompatibilists may welcome this conclusion and present it as an effective *reductio* of the compatibilist position. I argue, however, that incompatibilism, whether it takes the form of libertarianism or skepticism, encounters its own distinct difficulties when it comes to dealing with these broad concerns relating to fate and luck and the role they play in moral life. The conclusion that is drawn from all this is that although all the major parties in the free will debate reject (the very possibility of) free will pessimism, this is, nevertheless, the most truthful and accurate account of human agency and moral life. Critical compatibilism, I maintain, does not aim to solve the free will problem in terms that will satisfy the demands of the morality system. What we have is not a problem that needs to be solved but rather a troubling predicament that needs to be recognized and acknowledged.

References

Berofsky, Bernard. 2012. *Nature’s Challenge to Free Will*. Oxford: Oxford University Press.

— — —. 2017. “Classical Compatibilism.” In Meghan Griffith, Neal Levy, Kevin Timpe, eds. *Routledge Companion to Free Will*. New York & London: Routledge.

Dennett, Daniel. 1984. *Elbow Room: Varieties of Free Will Worth Wanting*. Oxford: Oxford University Press.

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

Honderich, Ted. 2002. *How Free Are You? The Determinism Problem*. 2nd ed. Oxford: Oxford University Press.

Kane, Robert. 1996. *The Significance of Free Will*. New York & Oxford: Oxford University Press.

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

McKenna, Michael & Paul Russell, eds. 2008. *Free Will and Reactive Attitudes*. Aldershot, UK: Ashgate Press.

p. xxiv Miles, James. 2015. *The Free Will Delusion: How We Settled for the Illusion of Morality*. Kibworth Beauchamp, Leics.: Matador.

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

Russell, Paul. 1995. *Freedom and Moral Sentiment: Strawson’s Way of Naturalizing Responsibility*. New York & Oxford: Oxford University Press.

Russell, Paul, 2015. “Hume’s ‘Lengthy Digression’: Free Will in the *Treatise*,” in *Hume’s Treatise: A Critical Guide*, edited by A. Butler and D. Ainslie (Cambridge: Cambridge University Press), 230–51.

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

Smith, Angie. 2007. “On Being and Holding Responsible.” *Journal of Ethics* 11: 465–84.

Sommers, Tamler. 2007. “The objective attitude,” *Philosophical Quarterly* 57: 321–41.

Sorabji, Richard. 1980. *Necessity, Cause and Blame*. London: Duckworth.

Strawson, P.F. (1962). “Freedom and Resentment.” Reprinted in Russell & Deery, 2013.

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

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

Williams, Bernard. 1965. “Morality and the emotions.” Reprinted in *Problems of the Self*. Cambridge: Cambridge University Press.

— — —. 1994. “Descartes and the Historiography of Philosophy.” Reprinted in *The Sense of the Past: Essays in the History of Philosophy*. Edited and introduced by Myles Burnyeat. Princeton: Princeton University Press.

— — —. 2000. “Philosophy as a Humanistic Discipline.” Reprinted in *Philosophy as a Humanistic Discipline*. Edited and introduced by A.W. Moore. Princeton: Princeton University Press.

— — —. 2002. “Why Philosophy Needs History.” Reprinted in *Essays and Reviews 1959–2002*. Foreword by Michael Wood. Princeton: Princeton University Press.

Notes

1. Relevant selections of these contributions can be found in Russell & Deery, 2013. See also the Introduction for a discussion and account of these developments, along with a brief survey of the key positions and strategies in the contemporary free will debate.
2. The naturalistic turn in recent years is, of course, itself something of a return to views articulated and defended by earlier figures, such as Hume. Contributions by Strawson (1962) and Williams (1965), among others, were especially influential in launching this process. For an account of the particular importance of Strawson's contribution in relation to this see the Introduction in McKenna & Russell (2008).
3. On this see, for example, Williams, 1994; Williams 2000; and Williams, 2002. Williams wrote several pieces concerning the relevance of history for philosophy and for ethics in particular. He also expressed skepticism about the "scientific" tendencies that are present, if not prevalent, in contemporary philosophy
4. Classical compatibilism is, of course, associated primarily with the central figures of the British empiricist tradition (viz. Hume, Mill, Russell, Schlick, Ayer et al.) but it still commands strong support and has influential defenders. See, e.g., Berofsky, 2012; and Berofsky, 2017.