**A.**

**The Philosophy of Free Will:**

*Selected Readings from the Contemporary Debate*

Paul Russell

&

Oisin Deery

Editors

February 2012

**Table of Contents**

**Introduction**

**I. The Free Will Problem – Real or illusory?**

1. Thomas Nagel - Moral Luck
2. Daniel Dennett - Please Don't Feed the Bugbears

**II. Naturalism Against Scepticism**

1. P. F. Strawson - Freedom and Resentment
2. Gary Watson - Responsibility and the Limits of Evil: Variations on a

Strawsonian Theme

**III. The Consequence Argument**

1. Peter van Inwagen - The Incompatibility of Free Will and Determinism
2. Dana Nelkin - The Consequence Argument and the Mind Argument

**IV. Responsibility and Alternative Possibilities**

1. Harry Frankfurt - Alternate Possibilities and Moral Responsibility
2. Michael Otsuka - Incompatibilism and the Avoidability of Blame
3. Kadri Vihvelin - Free Will Demystified: A Dispositional Account

**V. Libertarian Alternatives – Soft and Hard**

10. Robert Kane - Responsibility, Luck and Chance: Reflections on Free Will and Determinism

11. Randolph Clarke - Towards a Credible Agent-Causal Account of Free Will

13. Timothy O’Connor - Agent-Causal Power

**VI. Compatibilism: Hierarchical Theories and Manipulation Problems**

1. Harry Frankfurt - Freedom of the Will and the Concept of a Person

15. Richard Double - Puppeteers, Hypnotists, and Neurosurgeons

**VII. Compatibilism: Reason-Based Alternatives**

16.Susan Wolf - Sanity and the Metaphysics of Responsibility

17. John M. Fischer - My Compatibilism

**VIII. Autonomy and History**

18. John Christman - Autonomy and Personal History

19. Michael McKenna - Responsibility & Globally Manipulated Agents

**IX. Scepticism, Illusionism and Revisionism**

20. Galen Strawson - The Impossibility of Ultimate Moral Responsibility

21. Saul Smilansky - Free Will: From Nature to Illusion

22. Manuel Vargas - How To Solve the free Will Problem

**X. Optimism, Pessimism and their Modes**

23. Derk Pereboom - Optimistic Skepticism about Free Will

24. Paul Russell - Compatibilist-Fatalism

**XI. The Phenomenology of Agency and Experimental Philosophy**

25. Benjamin Libet - Do we have free will?

26. Eddy Nahmias, Stephen Morris, Thomas Nadelhoffer & Jason Turner

- The Phenomenology of Free Will

27. Shaun Nichols & Joshua Knobe

- Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions

**Bibliography**

**Index**

# *If having free will matters, it must be because not having free will would be awful, and there must be some grounds for doubting that we have it. What are we afraid of? We are afraid of not having free will. But what exactly are we afraid of? And why?*

# *--* Daniel Dennett*, Elbow Room*

It is often said that while philosophers may be good at asking questions their success in answering them is less impressive. These doubts may seem particularly apt in the case of the free will problem – an issue that has been debated for many centuries in one form or another. There is, nevertheless, good reason to look more favourably on the accomplishments of philosophy in light of recent developments in the free will debate. Although it may be said with some justice that the free will debate stalled and became stagnant for a long period of time, lasting well into the twentieth century, this cannot be accurately said of the debate over the past half century. Any student studying the core philosophical literature in period around the middle of the twentieth century would be immediately struck by the contrast in complexity and subtlety, as well as the variety, of positions now available to their contemporary counterpart. This is, moreover, not just a feature of one side or one party in the debate - it is true right across the board. All the major parties in the free will debate, judged by any reasonable philosophical standard, have had to revise and amend their core positions and strategies with a view to meeting the more sophisticated and nuanced positions and arguments advanced by their philosophical critics and opponents. If progress in philosophy is measured only in terms of securing consensus and agreement, or finding a view or position which all philosophers have come to accept and acknowledge, then the free will debate has not progressed a great deal, judged in these terms. However, if progress is to be measured by the philosophical quality of the positions offered, understood in terms of their flexibility and ingenuity in dealing with familiar objections and failings, as well as finding new ground on which to develop alternative positions and strategies that offer a way through or around these difficulties and objections, then considerable and significant progress has been made in this area during the past half-century. Indeed, the range and number of valuable and significant new initiatives and contributions is so extensive that it is proving hard for anthologies and collections to keep up to date and at the cutting edge of this debate. This volume has been gathered together with the aim of providing contemporary readers with a clearer view of the field as it has recently evolved, identifying particularly important issues and proposals that are currently under consideration.

Although it is our intention to allow the contributions in this volume to speak for themselves, it will be helpful, nevertheless, if we provide our readers with a brief overview of the background classical free will debate, along with a sketch of the central themes and issues of the contemporary debate that we will use to structure and arrange the contributions we have included in this volume.

***The Classical Free Will Debate***

Over a period of many centuries the problem of free will has taken a variety of forms and raised a number of interconnected issues and difficulties. The fundamental question at issue is the extent to which human agents are in control of their own lives and destinies. In different historical and social contexts various aspects of this problem have proved to be more salient than others. For many centuries the central preoccupation of philosophers concerned the religious difficulties involved in explaining how human freedom and moral responsibility could be rendered consistent with established assumptions about divine foreknowledge and predestination. In the early modern period the focus of attention switched from these theological aspects of the free will problem to issues arising from the advance of modern science and its metaphysical assumption. More specifically, when the concepts and categories of natural science were extended to include human thought and action, viewed as part of the seamless natural order of things, sceptical problems were generated relating to our self-image as free and responsible beings.

One familiar way of presenting these problems, as formulated in classical free will debate associated with early modern philosophy, is through the dilemma of determinism. The thesis of determinism may be broadly interpreted as the claim that everything that happens in the world, including human thought and action, is subject to causal laws. While the analysis of causal laws and causal relations may vary, it has been widely accepted that causation involves necessitation of effects by antecedent causal conditions. It follows from this that insofar as human action is caused it is thereby necessitated by causal antecedents in such a way that the action *must* follow, given the antecedent conditions. These assumptions are, according to some philosophers, *incompatible* with human freedom and moral responsibility.

One argument that classical incompatibilists have advanced in support of this claim concerns the issue of alternate possibilities and an agent’s ability to act otherwise. It is argued that moral responsibility requires the sort of freedom which ensures that the agent could have acted otherwise. If an agent’s actions are causally necessitated then there is no real open alternative available, given the actual conditions. It follows from this that determinism rules out the sort the sort of freedom required for moral responsibility. Classical compatibilists have rejected this line of argument and maintain, on the contrary, that freedom and moral responsibility are entirely consistent with determinism. The crucial distinction to get clear about here, they argue, is the difference between causal necessitation and coercion or compulsion of some kind. While it is true that freedom requires an absence of coercion and compulsion, causal necessity does not imply any form of coercion or compulsion. Conduct is compelled or coerced when it is *caused in a particular way*, when it is produced without or against the agent’s will and desires. However, when action is caused by the agent’s own will or desires then the agent can be said to act freely in the sense required for moral responsibility. Granted that the agent brings about the action through her own willings, the action is plainly attributable to her. Moreover, as long as action is produced by the agent’s own will or desires we are in a position to (causally) influence the agent’s future conduct by means of the motives of rewards and punishments. When an agent is free in this sense, these incentives will serve to guide and determine their conduct in socially desirable ways. Clearly, then, the only sort of freedom required for moral responsibility is a freedom to act according to the determination of our own will.

Classical incompatibilists reject this strategy and object, in particular, to its forward-looking, utilitarian understanding of moral responsibility. Children and animals, they argue, also act according to their own will and may also be influenced by rewards and punishments. They are, nevertheless, obviously not moral agents who can be properly held accountable for their actions. It is evident, therefore, that something crucial is missing from the classical compatibilist account of freedom and responsibility. According to incompatibilists, the crucial missing element is captured by the notion of *moral desert*, which requires that the agent is in some relevant sense the source or origin of her conduct. For an agent to be the originator or source of her conduct it will not suffice that she acts freely (i.e. without coercion, compulsion, etc.). It is also necessary that she was able to *choose* between genuinely open alternatives, and this requires the falsity of determinism and the absence of causal necessity. What is needed, on this analysis, is freedom to choose among alternatives possibilities in a way in which the outcome depends solely on the agent and not on prior causal conditions. What is required is not simply of free action but of *freedom of choice or free will*. Those incompatibilists who believe that these metaphysical requirements can be met or satisfied are known as *libertarians*.

One common objection to libertarianism, from a classical compatibilist perspective, is that there is a significant difference between an action being free and an action being simply uncaused. If an action lacks any cause, critics argue, then it is merely a chance event that cannot be attributed to any agent. In order for libertarians to avoid the difficulties associated with the horn of chance – i.e. the other side of the dilemma of determinism – it is crucial that they provide some account of free will that goes beyond the simple negative claim that our (free) actions are not causally necessitated. That is to say, libertarians need to be able to provide some metaphysical content to their notion of free will, or genuinely open choice, that does not collapse into mere chance or capriciousness. According to their critics, when libertarians try to meet this challenge their metaphysical commitments inevitably take the form of anti-naturalistic accounts that are obscure and impossible to integrate into the natural order of events in the world. The costs of adopting “spooky metaphysics” of this kind – such as non-empirical selves, special modes of causation, and so on - are both unnecessary and unacceptable. In sum, there is no available solution to the free will problem on this side of the dilemma of determinism.

The trajectory of the classical free will debate, given the difficulties encountered on both sides of the dilemma of determinism, threaten to lead us directly to the sceptical conclusion that the sort of freedom required for moral responsibility is impossible. On the side of compatibilism, the theory of freedom provided is too thin and frail to secure the sorts of distinctions that we need. A mere ability to act according to the determination of our own will, or to be subject to the influence of motives such as rewards and punishments, cannot serve as a credible foundation for moral life. On the other hand, when libertarians undertake to describe alternative metaphysical foundations that involve the absence of determinism and causal necessity their accounts are not only obscure in nature, they also encounter their own particular set of difficulties relating to chance and luck. The upshot of all this is that a general scepticism about human freedom and moral responsibility seems unavoidable. Free will, like the existence of God or the immortality of the soul, is a metaphysical illusion that we must now abandon.

It is a rather striking fact about the history of the classical free will debate that the radical sceptical conclusion has found relatively few defenders or adherents (in contrast, it may be said, with scepticism about the existence of God or the immortality of the soul). The explanation for this is that a radical scepticism of this kind seems both intellectually incredible and humanly impossible for us to live with or accept in practice. At one level this raises the concern that our core self-image as free and responsible agents is an indispensible part of our natural human make-up, and as such immune to any sceptical challenges presented on the basis of philosophical reflection (e.g. in much the same way as sceptical challenges relating to the existence of the external world or other minds may also fail to secure any sincere or consistent acceptance – however difficult the arguments advanced may be to refute). At another level, resistance to radical scepticism with respect to our self-image as free and responsible beings may reflect our *practical* interest in this matter. Viewing ourselves as beings incapable of genuine free, moral agency would disturb so many features and aspects of our lives that we value and care about – for example, relating to dignity, self-respect, creativity, and human emotions and relationships – that a deep pessimism about the human predicament would fall upon us and become unbearable of we were to sincerely embrace such an outlook or attitude to life. As beings who must live and act in the world, a gloomy pessimism of this sort is one that we will naturally and reasonably resist, whatever philosophical arguments may put to us by the sceptic. Nevertheless, be this as it may, these observations do not, in themselves, refute the sceptic, so much as invite us to *ignore* the sceptic – which is a different matter. As things stand with the classical debate, none of the available positions seem to be philosophically secure or comfortable. For a period of several centuries, stretching from the middle of the seventeenth century through to the middle of the twentieth century, these familiar arguments of compatibilists, libertarians and sceptics have been repeated and revised. The key battle-lines of the debate have remained largely unchanged, with little movement or variation in the strategies on offer and no sight of some credible solution on the horizon.

***The Contemporary Debate:***

The readings in this volume make clear that the free will debate is no longer stalled or stagnant. These readings have been selected and arranged with a view to highlighting the main contours of the current discussion by identifying the key issues and more significant contributions. The remarks that follow provide a brief description of the contents of the eleven sections of this volume and explain how they are structured and related to each other.

*I. The Free Will Problem – Real or illusory?*

Classical compatibilists have often argued that the free will problem is either merely “verbal” in nature or simply a “pseudo-problem”. The general force of these claims was to show that the obscure metaphysics of libertariaism is largely motivated by incompatibilist confusions and muddles that generate artificial and groundless worries and anxieties about the implications of determinism. In reply to this, incompatibilists maintain that classical compatibilism fails to appreciate the genuine difficulties that deterministic assumptions pose and that the sort of easy solutions compatibilists have offered fail to address the real and substantial difficulties we encounter here. The selections by Daniel Dennett and Thomas Nagel fall on either side of this dispute. According to Dennett, the classical free will problem is largely a product of “fearmongery” by philosophers. The worries and anxieties generated on this subject are produced by false and misleading analogies in respect of the implications of determinism for the human condition. Once these misleading analogies and “intuition pumps” are exposed and discredited, the free will problem will simply evaporate.

In contrast with this, Nagel aims to show just how deep and perplexing the free will problem is and the way in which it is systematically and intimately connected with many other fundamental problems of philosophy. Nagel approaches this task by way of considering various ways in which agents lack control over their conduct and character, leaving them vulnerable to the influence of fate and luck. The overall effect of these reflections on the limits of human agency is to erode our confidence that agents can ever be legitimately held morally responsible for anything that they do, as ultimately everything an agent does is a result of factors that she does not control. Nagel does not aim to solve this problem so much as to explain its roots in the split between our internal (subjective) experience of being agents and an external (objective) perspective which presents our actions as mere events that are part of the natural causal order. The intractable nature of the free will problem, Nagel maintains, has its source in this clash of perspectives with respect to human agency.

*II. Naturalism Against Scepticism*

Throughout much of the twentieth century, consistent with widely accepted views about the nature of philosophical methodology, it was assumed that the right approach for solving the free will problem was to provide a logical or conceptual analysis of the key (problematic) terms in this debate. With this in mind, much of the relevant literature was devoted to arguments about how to analyse different senses of “freedom”, “causation”, and their relations with “moral responsibility” and other similar notions. For the most part this was seen as an a priori or “armchair” investigation, requiring the special gifts and training of the philosopher to spot philosophical ambiguity or confusion. This way of thinking about the free will problem was challenged in a radical and innovative way by P.F. Strawson in his influential paper “Freedom and Resentment” (1962) – a contribution that has done much to stimulate the current debate. In this paper Strawson aims to find an acceptable middle-ground between classical compatibilists and their libertarian/incompatibilist opponents. Strawson invites us to begin our investigations, not with a priori analyses of the meanings of the key terms involved, but rather with a careful consideration of some familiar psychological facts about human nature and our social attitudes and practices. More specifically, his approach begins by drawing attention to our natural human “reactive attitudes” (i.e. moral sentiments broadly conceived), understood as basic human responses to the attitudes and intentions of people we are dealing with in social life. On the basis of an analysis of the rationale of excusing considerations Strawson goes on to argue that with respect to our reactive attitudes considerations of determinism do not engage or fall within the scope of any recognized excusing considerations. According to Strawson, moral responsibility has to be understood and interpreted in terms of this general psychological and normative framework or “web” of our reactive attitudes. It follows from this that determinism poses no threat to our commitment to moral responsibility understood in these terms. While classical compatibilists were right about this issue, their incompatibilist critics were also correct in claiming that “something vital” is missing from classical compatibilist accounts. The vital element that is missing, however, is not “contra-causal freedom” of the sort that libertarians are searching after but rather a proper appreciation of the role of reactive attitudes play in this sphere – something both sides of this debate have failed to recognize and give appropriate weight to. One particularly controversial feature of Strawson’s wider argument, is his claim that a proper appreciation of these psychological facts regarding the place and role of reactive attitudes in human life suggests that we are immune to the force of all sceptical arguments insofar as they aim to discredit and systematically dislodge our (natural) human commitment to the attitudes and practices associated with moral responsibility.

In the decades that have followed, Strawson’s approach to the free will problem through reactive attitudes or moral sentiments has generated a large volume of literature. One of the most engaging and insightful contributions to this literature is Gary Watson’s “Responsibility and the Limits of Evil”. Although Watson’s discussion is highly sympathetic, he raises several “hard questions” for Strawson’s (expressivist) theory of responsibility. One of these is how exactly we are to identify the boundaries of the moral community as constituted by those individuals who are appropriate targets of our reactive attitudes? What is needed here, Watson argues, is a more precise account of the nature of exempting conditions, whereby some individuals (e.g. the immature, the insane, etc.) are excluded as legitimate targets of our reactive attitudes. According to Watson, Strawson does not provide any full or adequate interpretation of this important aspect of the rationale of excuses. Watson presents a powerful example of the sort of difficulties we face by way of his detailed description of the disturbing case of Robert Harris, a convicted murder. Harris’s extreme evil, Watson suggests, makes it hard to decide whether or not he is really a proper target for reactive attitudes, since he lacks any shared framework of values that make for conditions of “intelligible moral address”. In relation to this case, Watson describes the tension that we experience when we reflect not only on Harris’s crimes but also on his horrific background of childhood abuse and deprivation. One general conclusion that Watson draws from his reflections on the case of Robert Harris is that, contrary to Strawson, it is not obvious that our reactive attitudes are immune to historical considerations in the way Strawson supposes. (Here there are obvious points of overlap with Nagel’s discussion of moral luck that merit further consideration.)

*III. The Consequence Argument*

While it is Strawson’s central aim in “Freedom and Resentment” to ally any concern that the thesis of determinism poses a threat to freedom and moral responsibility his arguments have not persuaded incompatibilists. A fundamental and long-standing incompatibilist objection concerns how, on the assumption of determinism, events in the distant past along with the laws of nature are related our actions and our ability to do otherwise. In his *An Essay on Free Will* (1983 ) Peter van Inwagen aims to articulate and present these concerns in a more rigorous form. He begins with an informal statement of his argument.

If determinism is true, then our acts are the consequences of the laws of nature and events in the remote past. But it is not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.

The basic intuition lying behind this argument is that if we cannot change something, then we cannot change anything that is a necessary consequence of it. For example, if it is true that the sun will explode in 2020 and no one has any choice about this, and also true that if the sun explodes all life on earth will die out and no one has any choice about this either, then it follows that no one has any choice about whether all life on earth will die out in 2020. In cases of this kind there is a “transfer of powerlessness” from one state of affairs to another. According to the consequence argument the thesis of determinism implies that everyone is powerless with respect to all our actions, insofar as they are understood to be the necessary consequences of prior conditions and laws that we cannot change and have no choice about.

Van Inwagen’s more formal presentation of this argument relies upon some more technical apparatus. He begins by introducing the expression “Np” – where “p” represents any sentence, and “Np” is read as, “p and no one has, or ever had, any choice about whether p”. To have a choice about p is to be able to render p false. Additionally, van Inwagen employs two inference rules. The first, Rule Alpha, states that if a proposition is logically necessary then no one has any choice about it (i.e. it cannot be rendered false). The second, Rule Beta, states that if no one has any choice about p, and no one has any choice about p logically implying q, then no one has any choice about q. Using this apparatus van Inwagen constructs a formal argument to show that if determinism is true, a proposition describing the state of the world in the remote past, in conjunction with a proposition describing the laws of nature, logically entails any true proposition, including all true propositions describing any actions we may perform. Granted we have no choice about the past and the laws of nature, and no choice about their consequences (i.e. what is logically implied by these propositions), it follows, by Rule Beta, that we have no choice about any of our actions (i.e. we cannot render false any of the propositions describing these actions). If this is the case we cannot do otherwise and, assuming free will requires the power to do otherwise, it follows that no one has free will.

Much of the debate is in response to van Inwagen’s statement of the consequence argument has turned on the validity of Rule Beta. In the reading that follows Dana Nelkin, following up on some of the critical literature on this topic, considers a serious problem facing libertarians who want to rely on Rule Beta and the Consequence Argument. The difficulty they face is that Rule Beta also lends support to an argument for the incompatibility of freedom and indeterminism – the influential (anti-libertarian) Mind argument, as it is known. The basic idea behind the Mind argument is that in an indeterministic world our actions are caused but not determined by prior states of mind of the agent. However, as ultimately no one has a choice about the causes of these states of mind, and no one has a choice about whether these states of mind will have a particular result or not, it follows that no one has a choice about the actions that they may cause. If this is the case, then adopting Rule Beta leads to skepticism about the very possibility of freedom understood as the ability to do otherwise. This is, of course, a conclusion that both compatibilists and libertarians will want to avoid.

Nelkin’s discussion of this issue examines various ways that compatibilists and libertarians have responded to the challenges presented by Rue Beta. Some compatibilists have argued that Rule Beta is false on the ground that it is vulnerable to counter examples based on concerns about agglomeration. In reply, libertarian incompatibilists have proposed a revised version of Rule Beta – Beta 2 – which, they argue, still serves the purposes of the Consequence argument but is not consistent with the Mind argument. This way the consequence Argument can be saved but the Mind Argument fails. In response to this Nelkin argues that the Mind argument can be reformulated using Beta 2 and, therefore, both arguments stand or fall together. She concludes by noting that while it may be true that we cannot ensure that we act other than we actually do, this may not be the freedom we care about. If we understand acting freely in terms of our states of mind causing our actions we can escape both incompatibilist and skeptical conclusions.

*IV. Responsibility and Alternative Possibilities*

If moral responsibility requires free will, and free will requires being able to do otherwise, then moral responsibility requires being able to do otherwise. This view is formulated by Harry Frankfurt as the Principle of Alternate Possibilities (PAP) and it is a view that was once shared by all parties in the free-will debate, including compatibilists. As we have noted, one particularly influential compatibilist strategy was to interpret the claim that we could have done otherwise as meaning that if the agent chose or willed to do otherwise he would have done so. This hypothetical analysis has not, however, convinced incompatibilist critics, who argue that we require a categorical analysis of this ability, one that does not suppose our choices or willings were different in the actual circumstances. Harry Frankfurt has suggested a way in which the difficulty relating to PAP may be side-stepped. He does this by describing a counter-example to PAP which establishes that to be responsible it is not necessary that an agent be able to do otherwise. The example he describes involves a neurosurgeon (Black) who wants

If free will is required for moral responsibility, then if free will requires being able to do otherwise, so too does responsibility. This view is formulated by Harry Frankfurt as the Principle of Alternate Possibilities (PAP). All parties to the free-will debate once shared this view, which is why compatibilists gave a compatibilist-friendly analysis of ability. Yet, as we saw in our discussion of the consequence argument, there are reasons for thinking that such an analysis fails. Frankfurt plugs the resulting gap in compatibilist theories by arguing that PAP is, in any case, false: as it turns out, we don’t need to be able to do otherwise to be morally responsible. Frankfurt argues for this conclusion by means of a counterexample case to PAP.

In Frankfurt’s original case, Black — a neurosurgeon — wants Jones to choose A. He is able to intervene to control Jones’ brain processes should Jones be about to choose B. Yet, preferring not to intervene, Black instead waits to see how Jones will choose on his own. Jones is unaware of Black’s presence. Frankfurt claims that Jones lacks alternative possibilities in this case. Still, if Jones chooses A on his own, it seems he is morally responsible for his choice, even though he had no alternative possibility should he have decided to do otherwise. After all, Black was ready to intervene to control Jones’ brain processes. So PAP is false. Thus, both traditional incompatibilism and classical compatibilism are false, since each relies on PAP.

Michael Otsuka’s paper, “Incompatibilism and the Avoidability of Blame”, comprises a libertarian response to Frankfurt’s paper. Whereas some libertarians deny that Frankfurt-style cases are genuine counterexamples to PAP, Otsuka accepts this verdict. Yet although Otsuka concedes that we should reject PAP, he thinks that the libertarian can replace this principle with what he calls “The Principle of Avoidable Blame”. According to this principle, “One is blameworthy for performing an act of a given type only if one could instead have behaved in a manner for which one would have been entirely blameless.” Otsuka argues that such a principle is immune to Frankfurt’s counterexample, which only works against PAP. For Otsuka, even if it is true that Jones could not have chosen B and thus had no alternative possibility at the moment of his choice (since Black would have intervened had Jones been about to choose B), it is still true that *if* Jones had been about to choose B and consequently Black had intervened to manipulate Jones’ brain processes in order to make him choose A, choosing A would have been an action for which Jones was entirely blameless. After all, Otsuka argues, both compatibilists and incompatibilists usually agree that being directly manipulated undermines blameworthiness.

In “Free Will Demystified: A Dispositional Account”, Kadri Vihvelin argues — from the *compatibilist’s* point of view — that Frankfurt’s argument against PAP fails. First, she argues that even though the classical compatibilist analysis of the ability to do otherwise fails (see our discussion of the consequence argument), there is a viable account of the ability to do otherwise available to the compatibilist. According to Vihvelin, an ability to act (or not to act, which is just to be able to act in another way) is analyzable along the following lines: an agent can do A at t1 (say, raise her hand at t1) just in case were she to choose to do A at t2, and her body stayed working normally and nothing interfered with her, she would A at t2. In other words, Vihvelin holds that “persons have abilities by having intrinsic properties that are the causal basis of the ability”. So Vihvelin thinks that an ability to act is a disposition, or a bundle of dispositions. And, as she points out, “no one denies that dispositions are compatible with determinism”. After all, even if determinism is true, glass is still fragile — i.e., it has the disposition to break if struck.

If this account is right, then Vihvelin thinks that compatibilists *shouldn’t* abandon PAP. After all, Frankfurt’s argument is supposed to work no matter how we understand the ability to do otherwise. The crucial step in his argument is that Black’s presence has the result that Jones is unable to do otherwise, since he lacks alternative possibilities. But on Vihvelin’s account of the ability to do otherwise, Jones *is* able to do otherwise, even with Black on the scene. So on Vihvelin’s compatibilist understanding of the ability to do otherwise, Frankfurt’s argument fails. The conclusion Vihvelin wants us to draw is that compatibilists should retain their classical position: agents can do otherwise even if determinism is true, and this ability is required for moral responsibility.[[1]](#endnote-1)

*V. Libertarian Alternatives — Soft and Hard*

Libertarians are incompatibilists who reject skepticism about freedom and moral responsibility. On their account, the possibility of and existence of human freedom and moral responsibility depends on indeterministic metaphysics. With regard to this, as we have noted, libertarians face the challenge of fending off the objection that placing human freedom and moral responsibility on the metaphysical foundations of indeterminism serves only to generate problems of chance and luck, which erodes and undermines rather than supports freedom and responsibility. Even if this is not the case, it may still be argued that indeterminism does not and cannot *enhance* or *expand* our freedom and responsibility. The basic problem for libertarians, therefore, is to explain how the absence of determinism contributes to human freedom and moral responsibility.

The traditional libertarian approach, reaching back to figures such as Thomas Reid and Immanuel Kant, has been some form of agent causation. Agent causation is understood to be distinct from event-causation and may also be regarded as the only true or real causation. Although the details vary, all forms of the traditional view hold that when an agent acts with free will the event that is directly agent caused has no event-cause. The two crucial elements in this account are the conception of an agent as a substance (in earlier versions this is generally taken to be an immaterial, indivisible, intelligent substance – or a *soul*) that is not reducible to a series of events or states, and also a conception of causation such that something that is not an event (i.e. a substance) can bring about an event without there being any antecedent sufficient conditions that necessitate that event. Famously, the difficulty here is to avoid obscure or unintelligible metaphysical commitments that render libertarianism incoherent or inconsistent with the naturalism of modern science.

The three papers in this section each suggest a different way in which contemporary libertarians may attempt to meet this challenge. The first way, adopted by Robert Kane, is simply to abandon agent-causation accounts in favour of an event-causation account. A key feature of Kane’s account is that he combines an (event) causal theory of action with the view that causes need not necessitate their effects. This enables the libertarian to construct a theory of freedom and moral responsibility which turns on the notion of “self-forming actions” whereby an agent is presented with plural reasons for action such that the agent is not necessitated to act on any one of the available reasons and could act otherwise, on another available reason, in the same circumstances. However, when the agent acts on one or other of the available reasons the agent still acts voluntarily, intentionally and for a reason, even though she could have done otherwise in the same circumstances. Moreover, whichever way she acts, Kane argues, is the result of her choice and efforts. According to Kane, this model allows libertarians to reject the criticism that indeterminism introduces unacceptable, responsibility eroding forms of luck into our moral life. Beyond this, Kane also argues that his event-causal model of free action and moral agency is consistent with and supported by recent developments in neuroscience (e.g. relating to neutral networks) along with our understanding of the brain as a parallel processor – and so libertarianism of this kind in no way depends on obscure or unscientific metaphysical assumptions.

Unlike Kane, Randolph Clarke does not entirely abandon agent-causation as a feature of his alternative account of libertarianism. Although he argues that agent-causation in its traditional or pure form is vulnerable to the charge of being mysterious if not incomprehensible, he nevertheless believes it can be resuscitated in another form, one that integrates agent causation with universal event causation. Whereas the traditional view maintains that the relevant event that is directly caused by the agent when she acts freely has no event cause, on Clarke’s “causal agent-causal view” this event is non-deterministically caused by prior events. On Clarke’s account, “the agent’s acting with free will consists (crucially but not wholly) in her actions being caused, in this way, by her and by her reasons.” The agent’s reasons are prior events that have a causal role to play but do not necessitate or determine the action. Crucially, the agent also has a role to play in determining which of the particular reasons available become the cause of her actions. When an agent acts freely, on this account, she exercises a power that influences which of the alternative actions left open by prior events will actually be performed. In this way, the causal agent-causal view “mixes event causation and agent cauasation in its account of the production of action.” Since the agent exercises a further causal power of this kind she is, “in a strict and literal sense an originator of her action.” This alternative account, Clarke maintains, avoids many of the difficulties of the traditional view and, in particular, it provides a more naturalistic view of free action while still providing a role for the causal activity of the agent herself.

In the third and final contribution in this section Timothy O’Connor seeks to provide a defence and explanation of the traditional agent-causal view. Contrary to the widely held view that agent-causal theories are incoherent or unintelligible, Clarke aims to show that we can come to a clearer understanding of agent-causal power and the metaphysics of free will on the basis of a more general account of the metaphysics of causal powers. O’Connor provides a “structural tendency account” of agent causation. First, he assumes a metaphysical picture according to which the world is populated by primitive qualities whose nature is irreducibly dispositional. Such qualities involve “tendencies to interact with other qualities in producing some effect, or some range of possible effects”. Importantly, such dispositionality need not be deterministic; there may be “objective probabilities less than one that a cause will produce its characteristic effect on a given occasion”. O‘Connor applies this metaphysics to agents. According to O’Connor, “[A]n adequate account of freedom requires ... a notion of a distinctive variety of causal power, one which tradition dubs ‘agent-causal power’”. Of course, O’Connor doesn‘t insist that any agent *actually* possesses such a power, only that some possible agents (perhaps us) *could* possess it. O’Connor’s key requirements are that the agent’s power be ontologically both *emergent* and *basic*: it must confer on the system that possesses it the power non-redundantly to contribute to that system’s overall causal power. O’Connor thinks we can make sense of such emergent powers, “ones that are at once causally dependent on microphysically-based structural states and yet ontologically primitive”. The sort of power he envisages here would confer upon any agent possessing it an ability to *cause an intention*, or — as he puts it — to cause “an event which is the coming to be of a state of an intention”. The idea here is that reasons *causally structure* the agent-causal power by conferring continually varying objective probabilities on the agent’s causing certain intentions. But the forming of intentions is (crucially) goal-directed. By acting *on* a given reason, an agent may also act *for* that reason, when she causes an intention to act for the purpose of realizing a goal, and she believes that so acting will help her to achieve it.

*VI. Compatibilism: Hierarchical Theories and Manipulation Problems*

According to standard classical compatibilist accounts, the freedom required for moral responsibility is simply a freedom to do as we want or please. There is no metaphysical freedom of will lying behind this. This account, as we have noted, is vulnerable to a number of obvious objections. The most important of these being that it fails to distinguish responsible agents from children, the insane and animals – all of whom may enjoy freedom to act according to their desires. A plausible and satisfying account of freedom and responsibility requires something deeper than the classical compatibilist can provide us with. Harry Frankfurt aims to supply missing depth by means of his alternative hierarchical account. Frankfurt’s hierarchical account turns on several key distinctions. In the first place we need to distinguish between “persons”, who are able to form preferences about the structure of their will, from “wantons”, who lacks any such capacity. Persons, unlike wantons, are able to form second-order volitions about which of their first-order desires will become their will or be effective in leading to action. Wanton lack this capacity to reflect critically on their own will and for this reason freedom of the will, as opposed to freedom of action, is not a problem or them. In contrast with the wanton, persons may or may not reflectively identify with and endorse their own will. When a person’s will is not in conformity with his second-order preferences then, Frankfurt suggests, we may say it is not truly his own. The apparatus of second-order reflection allows us to speak of a person having a “real self”, as identified through their second-order preferences about the structure of their will. What is crucial for moral responsibility, on this view, is that the agent reflectively identifies with his own will. A person who has freedom of will is one who can reliably secure conformity of his will to his second-order volitions.

The most obvious objection to hierarchical or real self views of the kind that Frankfurt proposes is that a manipulator or “puppeteer” could insert or implant second-order preferences of the relevant kind, satisfying the conditions of responsibility or free agency as described by the hierarchical model. Intuitively such an agent is neither free nor responsible. The model cannot, therefore, critics maintain, provide us with a way of identifying sufficient conditions for free or responsible decisions. Richard Double suggests two possible lines of reply. First, granted that in ordinary cases of determinism we are not controlled by nature (since nature is not another agent), we may simply add a further condition that “free decisions require a normal etiology that excludes being brought about by external agents.” Second, we may also identify a set of characteristics (i.e. “autonomy variables”) that ensure that our decisions are ours irrespective of whether the causal that produces these decisions is determined or even controlled by a puppeteer. According to Double, as long as the agent has self-knowledge, and her states of mind are subject to critical self-assessment that effectively govern the agent’s will, then she has rational control, which is all that is needed with respect to free will.

*VII. Compatibilism: Reason-Based Alternatives*

Although Frankfurt maintains that the formation of second-order volitions requires rational capacities, he places no restriction on the causes or source of these preferences. Our second-order volitions (or even higher-order preferences) may be capricious and lacking any rational basis. What matters, for the purpose responsible agency, is simply that the person concerned *decisively* identifies himself with the relevant first-order desires. Double’s elaboration and revision of this model by way of the suggested “autonomy variables”, demands tighter control of our higher-order preferences by our rational capacities. From this perspective, what really matters for the purpose of freedom and responsibility, is not a hierarchical structure to our will as such, but rather that our will is subject to rational control. The papers by Susan Wolf and John Fischer present more elaborate versions of this general model. What is crucial to their proposals is that responsible agency requires some form of rational self-control that may be satisfied within the limits of compatibilist commitments.

Wolf argues that hierarchical views such as Frankfurt’s (what she refers to as broadly the “deep-self view”) correctly identify a necessary condition for responsibility but fails to provide a sufficient condition of responsibility, since it is not evident that even our deepest selves are up to us. While some may insist that what is required for responsibility is some form of self-creation, Wolf argues that all that is actually required, in addition to the original model, is that our deep-self be *sane*. In its essentials, what sanity requires is that we are able to understand and appreciate right and wrong, and to change our character and our actions accordingly. What the sane deep-self view presupposes is that responsible agents are able to *correct* or *improve* themselves. Individuals who come from deprived childhoods or are victims of corrupt societies and upbringings cannot help but be mistaken about their values and for this reason, Wolf maintains, we do not blame them for their actions that flow from their values.

John Fischer aims to combine the best features of incompatibilism and compatibilism, arriving at a position he calls “semi-compatibilism”. On the incompatibilist side of things, Fischer finds the Consequence Argument “highly plausible” and is disposed to accept its conclusion that determinism rules out freedom understood in terms of a capacity to select among open alternatives. On the other hand, Fischer is also persuaded by Frankfurt-style examples that suggest that we do not require alternative possibilities (or any freedom of this kind) for moral responsibility. He makes his case for compatibilism with respect to moral responsibility primarily on the basis of a distinction between two kinds of control. What he describes as “regulative control”, which requires alternate possibilities and the freedom to choose otherwise, is not required for moral responsibility. What is required is simply “guidance control”, where this is understood as effectively governing your choices and actions by your own reasons (e.g. as when you are in control when driving a car). As long as the agent has this kind of control operating in the actual-sequence, the absence of regulative control does not undermine the agent’s responsibility (as we find in Frankfurt-style cases). Fischer goes on to provide a general account of what is required for guidance control and moral responsibility. His account requires, first, that the agent can recognize reasons and, in particular, recognize moral reasons. Secondly, it also requires that the underlying “mechanism” or process of practical reasoning involved is in a relevant way the *agent’s own*. Ownership of this kind is explained by Fischer as a matter of “taking responsibility”, whereby the agent comes to see herself in a certain way – in a way that involves, among other things, seeing herself as a fair target of reactive attitudes based upon how she exercises her agency. When these two conditions are satisfied then our self-image as (robustly) responsible moral agents will be secured against any threat related to the truth of determinism.

*VIII. Autonomy and History*

Both hierarchical and reason-responsive compatibilist theories have done much to address some basic weaknesses and shortcomings in classical compatibilist accounts. This includes efforts to provide greater “depth” and complexity for compatibilists in a manner that allows it to draw several key distinctions that earlier, cruder views could not make. Having said this, however, a familiar and fundamental objection of incompatibilism persists: the worry that the sorts of conditions required by hierarchical or reasons-responsive theories could be satisfied consistent with the fact that the agents concerned have their (structured) wills or set of values and preferences implanted or manipulated by another agent (e.g. an evil neuro-surgeon, the Devil, etc.) This general problem, as discussed by Richard Double and John Fischer, raises worries about “ownership” or “autonomy” – where autonomy is understood in terms of an agent’s ability to be truly “self-governing” and motivated by desires and preference that are properly her own or “authentic” (i.e. not “alien” or “external”). The question arises, therefore, how compatibilists can meet this challenge. The “soft compatibilist” response is to agree that conditions of manipulation or implantation, as described, are indeed inconsistent with freedom and responsibility but that compatibilists have a principled way of dealing with cases of this kind by appealing to relevant “historical” considerations.

How can agent autonomy be protected or preserved by way of relevant historical considerations? John Christman presents an important statement of this approach, one that has been further developed and refined in recent philosophical literature. What Christman suggests is that we should “focus on the manner which the agent *came to have* a set of desires rather than her attitude towards the desires at any one time.” It is, he argues, the process of desire formation that is relevant to understanding the conditions of autonomy, and this must involve the agent’s *participation* in the formation of her preferences. More specifically, autonomy is satisfied only when an agent is aware of changes and developments of her character and is in a position to *resist* such changes on the basis of her self-understanding. These requirements, Christman maintains, serve to exclude cases where a person is manipulated or subject to social conditioning that undermines her reflective capacities. In general, the relevance of history for assessing issues of autonomy and ownership is that it aims to provide a basis for identifying deviant causal origins of an agent’s desires and preferences which undermine or erode freedom and responsibility. (One especially significant feature of Christman’s account of autonomy is that it is relevant to problems of political as well as moral freedom.)

In contrast with Christman’s approach, Michael McKenna aims to provide a “modest defence” of a nonhistorical compatibilist response to objections based on global manipulation (i.e. consistent with Frankfurt’s “tough stand” on manipulation cases). McKenna’s “hard compatibilist” strategy begins by assessing the case for (soft) historical compatibilism, giving particular attention to Alfred Mele’s influential efforts to support the conclusion that at least one necessary condition of “compatibilist-friendly agential structure” must be an historical one. McKenna argues that manipulation examples can be extended to cover historical conditions (i.e. relating to the way an agent’s values and attitudes have been produced) whereby even historical compatibilists, such as Mele, must accept that a manipulated agent may nevertheless be free and morally responsible for what she does. With this in mind, McKenna goes on to explain why we are reluctant to accept that a manipulated agent may be free and responsible. Two considerations are especially important according to McKenna’s hard compatibilist analysis. First, it is important to recognize that even if an agent who has been globally manipulated is free and responsible for the actions that flow from her values and attitudes, unlike a non-manipulated agent she is not responsible for coming to be the sort of person she is. With respect to this dimension of her agency there is a significant gap between the globally manipulated and non-manipulated agent which the hard compatibilist can and should recognize. Secondly, it is consistent with the hard, nonhistoricial compatibilist outlook to acknowledge that in at least some cases the globally manipulated agent may be wronged and have her rights violated by the process of global manipulation, when another agent decides what sort of a person she will be. Recognizing these points and the distinctions they involve helps us to account for any intuitive resistance we may have to a “tough” or “hardline” stance on this issue.

*IX. Scepticism, Illusionism and Revisionism*

Whatever their differences with respect to the issue of determinism, both compatibilists and libertarians believe that freedom and moral responsibility can be provided with philosophically sound foundations. In contrast with this, the sceptic about free will believes that no secure foundations can be provided. One version of this view is “hard determinism”; the claim that determinism is true and that compatibilism cannot provide an adequate or satisfactory theory of freedom and moral responsibility. An even stronger version of this view, however, is that whether determinism or indeterminism is true, it is impossible to provide any credible theory of freedom of the kind required for (true or ultimate) moral responsibility. This strong version of the sceptical position is defended by Galen Straswon in “The Impossibility of Ultimate Moral responsibility”. The key to Strawson’s sceptical position is what he calls the “Basic Argument”. Strawson provides several versions of this argument, each giving various degrees of detail and complexity. The most simple version, slightly amended, has the following three steps:

1. Nothing can be cause of itself (i.e. *causa sui*).

2. True moral responsibility requires that the agent is cause of himself, at least in certain crucial mental respects.

3. Therefore, true moral responsibility is impossible.

Strawson, as noted, refines and explains each one of the core premises of this argument, and defends them from various objections from both the compatibilist and libertarians perspectives.

A crucial issue of interpretation with respect to the Basic Argument is *how sceptical* it is intended to be. In particular, is the argument sceptical about moral responsibility *tout court* or just *ultimate* moral responsibility (i.e. considered as *one interpretation* of moral responsibility)? On the stronger reading, Strawson’s reference to ultimate moral responsibility as “true” or “real” may be taken to imply that any other accounts of moral responsibility fall short of the relevant standard of adequacy (e.g. they are shallow, superficial, etc.). The notion of “true moral responsibility” is one that involves a commitment to a certain established assumptions about agency and retributive justice. These are assumptions that are not only ingrained in our Western (Judeo-Christian) ethical outlook, they are also part of universal human experience of the phenomenology of choice and action. In contrast with this, a weaker reading may take the Basic Argument to be more limited in its objectives, holding that while “true moral responsibility” is impossible other (viable) forms may replace it without any catastrophic loss in our self-image as moral agents. (A weaker reading of this kind, limited in its sceptical ambitions, would bring the Basic Argument closer to the ground occupied by “revisionists” such as Manuel Vargas, as described below.)

A related but distinct position, building along the lines suggested by the scepticism of the basic argument, is presented by Saul Smilansky in “Free Will: From Nature to Illusion”. Smilansky’s analysis turns on three major questions. The first is whether libertarian free will actually exists. To which Smilansky answers in the negative. The second is whether moral responsibility and related notions are compatible with determinism, to which he gives a more qualified answer. While Smilansky acknowledges the “partial validity of compatibilism” with respect to a number of important distinctions that it identifies, he nevertheless finds it “shallow” on the grounds that it is unable to account for morally significant features of our lives. These morally significant features include, Smilansky argues, core values such as justice and desert as well as self-respect. Smilansky pays particular attention to the aims of (P.F. Strawson”s) “naturalism”, understood as an attempt to deflect and deflate the sceptical threat to our common sense attitudes and practices in this sphere. His basic objection to the naturalist project, as he interprets it, is that the reactive attitudes and emotions that it appeals to themselves depend on a belief in libertarian free will which cannot be defended light of critical reflection.

With these “pessimistic” answers to his first two questions in place, Smilansky goes on to consider the consequences and implications of the conclusions that he has reached. Much of his answer to this question consists or reviewing the extent and costs of our inability to defend any form of libertarian free will against skeptical critique. Smilansky suggests that when we consider the matter from “the ultimate perspective” we will find things are very bleak indeed. Compatibilist distinctions will provide only a small measure of support for our deepest values relating to our sense of self-worth and achievement, remorse and integrity, and so on. This leads us to the “grim” and “disheartening” conclusion that there is “no real substitute for the framework of achievement, desert and value based on free action.” The only possible escape from this “darkness” is to take the path offered by *Ilusionism*.

The final sections of Smilansky’s paper outline what the Illusionist path involves and what justifies it. Smilansky claims that illusory beliefs concerning free will are already in place and that they play a positive and necessary role in human life. (Here we may, of course, compare parallels with illusory beliefs about God or future state.) Although our belief that we possess libertarian free will is false we have overwhelming practical reason, Smilansky suggests, to do what we can to retain and support this belief For example, may do this by encouraging required form of self-deception and wishful thinking, as well as by discouraging all forms of critical reflection that are likely to undermine it. Without illusion of this kind, it is argued, not only will our sense of self-respect and our deeper moral values of justice be damaged, we will not even be able to effectively maintain compatibilist distinctions that we also need for ordinary moral life and practice. Smilansky concludes his defence of Illusionism by claiming “that our priority should be to live with the assumption of libertarian free will although there is no basis for this other than our very need to live with this assumption.”

In “How to Solve the Problem of Free Will” Manuel Vargas articulates an alternative way of approaching this issue. Like Galen Strawson and Saul Smilansky, Vargas argues that both the standard solutions, as advanced by libertarians and compatibilists, are not satisfying. While libertarianism captures important aspects of our everyday thinking on this subject, it also involves problematic metaphysics and makes claims that are epistemologically impossible to ever verify. When we turn to compatibilism familiar difficulties arise here as well. In particular, according to Vargas, compatibilist accounts of free will are plausible only to the extent that they ignore the very problematic commitments “that gave us a problem that needs solving”. Whereas Galen Strawson’s basic argument encourages the thought that we must “eliminate” or abandon responsibility in these circumstances, and Smilansky suggests that we should embrace illusion, Vargas charts another course.

Vargas’s revisionist approach turns on a fundamental distinction between *diagnostic* and *prescriptive* theorizing. A diagnostic account is one that aims to describe our ordinary, everyday thinking or established understanding of free will. In contrast with this, a revisionary account provides an interpretation of how we ought to think about these matters. Clearly it is possible for a gap to open up between these two distinct concerns, given that it may turn out, after philosophical reflection, that free will is not the sort of thing that we have hitherto supposed it to be. What we are looking for, from the prescriptive point of view, is a notion of responsible agency and a justifiable system of praising and blaming that “can adequately function without an incompatibilist metaphysics”. This is all we need, Vargas maintains, to be revisionists and avoid the sceptical conclusion that we must “eliminate” or abandon our commitment to the attitudes and practices associated with moral responsibility. The goal of such a revisionist theory is to “identify features of agency that can play an appropriate role in explaining the justification of our practices of praising and blaming and that helps explain how familiar judgments and attitudes about freedom and responsibility make sense in a framework that minimizes ad hoc metaphysical commitments”.

What, then, could justify praise and blame, when they are not founded upon libertarian metaphysics? Vargas suggests that praise and blame can be justified in terms of the role that they play in fostering those forms of agency that involve our ability to respond to moral considerations. A notion of free will developed along these lines is compatible with what we know about our agency and the world. Although it may require us to abandon that part of our self-conception that presents us as capable of initiating causal chains, it nevertheless serves the purpose of making sense of our moral lives and leading us away from “responsibility nihilism”. Understood this way, revisionism claims to find a way of acknowledging incompatibilist concerns at the diagnostic level, while at the same time developing a satisfactory compatibilist account of free will at the prescriptive level. In taking this approach we can avoid the perils and difficulties of both scepticism and illusionism.

*X. Optimism, Pessimism and their Modes*

Throughout the free will literature there is a close relationship between the alternative positions on this subject and metaphysical attitudes of optimism and pessimism. More specifically, the shared ambition of libertarians and compatibilists to somehow justify or vindicate free will and moral responsibility is generally associated with an optimistic outlook that holds that we lack those powers and capacities relating to agency and moral life that we most care about. On the other hand, the sceptical position is closely allied with a pessimistic outlook which claims to unmask our illusions relating to our self-image as genuine moral agents – a process which is understood to disturbing and damaging to our aspirations to view ourselves beings who uniquely govern nature without being governed by it. This simple dichotomy between free will optimists and sceptical pessimists is challenged in Derk Pereboom’s “Optimistic Skepticism about Free Will”.

Pereboom defends a position that he describes as a position that he describes as “Hard Incompatibilism”. There are two key elements to this position: first, it defends scepticism about free will and moral responsibility and, second, it argues that “a conception of life without this type of free will would not be devastating to morality or to our sense of meaning in life, and in certain respects it may even be beneficial”. The form of free will that Pereboom denies has any existence is the kind that is required for moral responsibility, where this is understood as involving being *deserving* of praise and blame. Pereboom directs his sceptical arguments at both compatibilist and libertarian accounts of free will. With regard to comaptibilism, Pereboom opts for “source” as opposed to “leeway” incompatibilism. Pereboom’s source incompatibilism is based on a version of the manipulation argument, whereby compatibilists are challenged to draw a principled distinction between four cases, beginning with a case of direct manipulation and extending down to a case of simple determinism by causal processes with no manipulation involved. In all these cases, Pereboom argues, the source of agent’s action can be traced back to factors beyond his control. Even though all relevant compatibilist conditions of responsibility are met in these cases, the agent is not intuitively responsible and there is no principled distinction to be drawn between the cases involving manipulation and the case of simple causal determinism. At the same time, Pereboom also rejects any effort to defend incompatibilism on the basis of considerations of alternative possibilities because he believes it is possible to construct a plausible Frankfurt-style case where an agent has no alternative possibilities but is nevertheless intuitively morally responsible. With regard to libertarianism, Pereboom argues that event-causal libertarianism (e.g. as advanced by Kane) is vulnerable to the “luck objection”, insofar as the agent lacks any control over whether a given decision occurs or not. On the other hand, agent-causal libertarian theories, although not vulnerable to the luck objection, are not credible. More specifically, while theories of this kind may be coherent they require us to accept that there is an incredible coincidence between what our best physical theories lead us to expect and what results from agent-caused free choices.

With the sceptical side of his argument in place, Pereboom then turns to the constructive side of the hard incompatibilist agenda. Contrary to the view that scepticism about free will and moral responsibility would be impossible to live with or accept in practice, Pereboom maintains that most of what we care about could be preserved and maintained. He begins by arguing that our basic moral distinctions between right and wrong could be preserved, as could the importance we attach to moral education and reform by means of “admonition and encouragement””. Similarly, although we would have to give up a retributivist justification for criminal punishment, we could still justify punishment for criminal behaviour in terms of education and moral improvement. We may also justifiably “quarantine” those who we deem criminally dangerous, even though these individuals are not responsible for being dangerous to others. Going beyond the boundaries of morality and law, Pereboom argues that there is no basis supposing that our lives would lose meaning or purpose if we accept the truth of hard incompatibilism. Nor do we have reason to suppose that important interpersonal relations would be seriously damaged or threatened. While some of our reactive attitudes would need to be abandoned (e.g. resentment and indignation) others – or “analogues” of them – would survive this alteration in our self-image. This includes, Pereboom argues, important forms of love that do not depend or rely upon the assumption that the object of these emotions is in some way morally deserving of them. Pereboom concludes that while there is a range of (reactive) emotions that may be lost or unavailable to us if we became convinced that we do not have the sort of free will required for moral responsibility this may be, on balance, a good thing.

It is generally accepted by both compatibilist and incompatibilists that responsibility and fate exclude each other – in circumstances where an agent is subject to fate she cannot be responsible for her conduct. However, although the two opposing parties are agreed on this point they understand fate, and its relation to determinism, in very different terms. Compatibilists standardly argue that fatalistic circumstances should be interpreted in terms of the causal inefficaciousness of an agent’s deliberations and actions. An agent is fated when what happens does not depend on what the agent decides or does. The thesis of determinism does not imply that all our actions are fated in this sense. Although fatalism of this kind is incompatible with moral responsibility, there is no threat to moral responsibility on the ground that determinism implies universal fatalism. In contrast with this, incompatibilists understand fate in less restricted terms, taking it to include, in particular, our concern that an agent is not the real or true originator or source of her conduct. According to incompatibilsts, fatalism of this kind is implied by the thesis of determinism and, in consequence of this, determinism implies that all our actions are fated and so no one is morally responsible.

In “Compatibilist-Fatalism” Paul Russell challenges the orthodox compatibilist position on this issue and argues that a plausible compatibilism requires a richer conception of fatalistic concern, one that recognizes the legitimacy of (pessimistic) concerns about the origination of character and conduct. The position generated by these compatibilist concessions to incompatibilism is described as “compatibilist fatalism” and involves two key claims. The first is the claim that determinism is compatible with moral responsibility and the second is that determinism implies conditions of universal fatalism, which serves as a genuine basis for pessimistic concern. With respect to the first claim, all compatibilists will agree that indeterministic metaphysics is not a necessary condition of moral responsibility and that a suitably “deep” account of moral responsibility can be provided within the constraints of compatibilist commitments. This is not, however, the focus of concern in this paper. The specific challenge for compatibilist-fatalism is to explain or account for the basis of pessimistic concern about origination where this does not involve worries about moral responsibility.

The independent source of pessimistic concern, Russell argues, should be understood in terms of the (Spinozistic) metaphor of our “sovereignty” over nature. More specifically, agents subject to origination-fate cannot conceive of themselves as (actively) ordering nature without being (passively) subject to its laws. Our self-image as a possessing God-like powers to initiate a causal series, and being something more than complex and sophisticated causal intermediaries, cannot be sustained in circumstances where origination-fate holds. While we may aspire to be able to intervene in the flow of natural events – and interrupt and redirect it – no compatibilist account of human agency can secure this view of ourselves. So considered, the exercise of our (moral) agency is itself, in the final analysis, subject external or alien causal influences. Any form of compatibilism that seeks to evade these troubling or disconcerting features of origination-fatalism, as found in the thesis of determinism, must be judged shallow and superficial (a charge that certainly applies to the orthodox attempt to restrict all worries about fate to issues of contribution and not origination). A plausible compatibilism, Russell maintains, must accept the relevance and legitimacy of pessimistic worries about origination, even if these worries should not be interpreted, as incompatibilists would suggest, in terms of scepticism about moral responsibility. In this way, a sensible compatibilism must acknowledge these pessimistic implications of determinism as this concerns the finite and limited nature of human agents even in circumstances where their powers of rational self-control and moral freedom are operating and exercised. (Compatibilist-fatalism may be understood as taking a view that is the opposite of Pereboom’s “hard incompatibilism”, insofar as the latter combines incomaptibilism with a broadly optimistic attitude to the implications of determinism for the human predicament.)

*XI. The Phenomenology of Agency and Experimental Philosophy*

Various experimental methods have recently been brought to bear on the free-will problem. For instance, in “Do we have free will?” neuroscientist Benjamin Libet describes the results of experiments he ran that show that a spike in cerebral neural activity — what he dubs the “readiness potential”, or “RP” for short — occurs up to a third of a second before people become consciously aware of having “an urge” to act. In his experiment, subjects were asked to flick their wrist whenever they wanted, while looking at an oscilloscope “clock” comprised of a rapidly revolving dot of light on a dial. Subjects were asked to report the position of this dot when they first became aware of forming an urge to flick their wrist. Using this method, Libet was able to measure to within a few milliseconds’ accuracy the onset of the urge to act. Meanwhile, an EEG machine recorded brain activity and an EMG machine measured muscular activity in subjects’ arms when they flicked their wrist. Libet’s results show that subjects’ reported awareness of feeling an urge to move their wrists occurred 200 ms. before the muscular activation recorded on the EMG. But the interesting result was that the RP measured by the EEG occurred earlier than this — 550 ms. before the muscle activation and 350 ms. before subjects reported feeling the urge to move their wrist. In addition, such RPs were not recorded when subjects’ didn’t report feeling an urge to move their wrist or in cases where subjects subsequently vetoed such an urge. Libet concludes that conscious control of our actions is limited to *vetoing* what the brain has already “decided” to do. A subject’s flicking her wrist is caused not by her conscious urge, but by the preceding RP. (For criticisms of this experiment see e.g., Mele 1995, 2009 and the other suggested readings.)

Complementing the work being done on free will by social psychologists and neuroscientists, some pioneering philosophers have lately “gone empirical” by rolling up their sleeves and actually conducting empirical studies themselves. These philosophers now comprise a growing movement called *experimental philosophy*, or X-Phi, and much of the early work done by these philosophers has focussed on free will.

Practitioners of X-Phi employ the methods of experimental psychology to probe folk intuitions about concepts like that of free will. Their aims include showing that a given view is *more* intuitive than another, or developing error theories to explain away either a majority or a minority intuition by uncovering the mechanisms that generate it and showing these to be <http://www.guardian.co.uk/world/2012/feb/14/is-religion-really-under-threat>unreliable. Indeed, the primary focus of much X-Phi is on uncovering just these cognitive mechanisms.

Taking the last two readings in this section in reverse order, in “Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions”, Shaun Nichols and Joshua Knobe investigate the psychological processes that generate competing compatibilist and incompatibilist judgments about responsibility. Their hypothesis is that people hold an incompatibilist theory of moral responsibility but possess psychological mechanisms that nonetheless produce compatibilist judgments in certain cases. By conducting survey-type experiments, Nichols and Knobe show that in *abstract* conditions, which were designed to elicit theoretical cognition, subjects tended to be incompatibilists about moral responsibility. Yet in *concrete* conditions, which were designed to elicit an affective response, subjects tended to be more strongly compatibilist. One of Nichols and Knobe’s central claims is that the affective response induces an *affective performance error* — in other words, such responses interfere with subjects’ ability to reason correctly. In this case, the error is just that they are unable to apply their incompatibilist theory of moral responsibility. Nichols and Knobe also show that whereas subjects’ judgments in *high-affect* conditions are compatibilist their judgments in *low-affect* conditions are incompatibilist, which further supports the hypothesis that affect plays a role in compatibilist judgments.

In “The Phenomenology of Free Will”, Nahmias and colleagues investigate how people describe their experience of agency (what they call the *phenomenology of free will*). In one study, subjects were asked about their current experience in an imagined case of decision-making. They were given descriptions of three apartments and were asked to choose an apartment as though they were going to live there for the next year. Subjects were asked to verbalize any thoughts that came to mind while they deliberated. The results indicated that what was most salient for subjects’ were the features of the apartments. In particular, they “did not mention anything suggestive of a self that determines the outcome of an otherwise undetermined choice”. Nahmias and colleagues conclude that this “supports the compatibilist description of the phenomenology more than the libertarian description”. In another experiment, compatibilism was pitted against incompatibilism more directly. Subjects in this experiment were asked to imagine an experience of making a difficult choice, and then to select which of two different (compatibilist and incompatibilist) descriptions of the ability to do otherwise best described their imagined experience. The result was that a majority of subjects selected the compatibilist description.

1. [↑](#endnote-ref-1)