Free Will, Determinism, and Moral Responsibility

Frank Arthurs

Thesis submitted for the degree of PhD in Philosophy

Department of Philosophy
University of Sheffield
March 2014

PREFACE

Since long before I began reading any academic literature on philosophy, I have been interested in the question of how we could have the free will required for moral responsibility. More recently, however, the key question has at times seemed to be not so much *how* we could have this kind of free will, but more fundamentally *whether* we have it. Before becoming exposed to (some might say corrupted by) the work of professional philosophers on this subject, it had not so much as crossed my mind that the possibility of possessing the free will required for moral responsibility might be called into question. I took it as read that libertarianism—with its belief that, in a given set of circumstances an agent is able to take more than one possible course of action—was the correct theory. However, upon sustained exposure to the seductive arguments of compatibilist philosophers, urging that the kind of 'open' futures libertarians posit are not necessary for free will, I saw my pre-philosophical intuitions begin to founder. When to these compatibilist arguments was added the influence of Spinoza and his remorselessly logical case for necessitarianism, there was nowhere left for my libertarian leanings to hide. In spite of myself, my Damascene conversion to determinism was effected.

Only then, once I had disavowed libertarianism, did the issue of whether we could have the free will required for moral responsibility really come into play. During my undergraduate career, I continued to wrestle with a dilemma that seemed to me to have no satisfactory solution. On one horn of this dilemma was compatibilism, with its various plausible-sounding but ultimately (I felt) specious explanations of how determinism posed no problem for our free will and moral responsibility. On the other horn was hard determinism, which, while it seemed more intellectually satisfying, demanded the impossible in the form of a repudiation of belief in moral responsibility. The dilemma remained disquietingly unresolved as I embarked on an MA thesis in Buddhist philosophy. Once again, my focus remained on questions of free will, determinism, and moral responsibility, although the task at this point was as much hermeneutical as it was purely philosophical.

The present PhD thesis, then, represents an opportunity to return to the purely philosophical task of resolving the dilemma I unwittingly found myself caught on the horns of, and marks the culmination of my thinking so far on this most absorbing, contentious, and intractable of philosophical issues. Still, it is merely that: the culmination of my thinking *so far*. I am conscious that it is a long way from being the finished product as it stands. In any case, even if I never revisit this topic in writing—which seems a welcome prospect after having just written the 77,777 words contained here—these certainly do not represent my final thoughts on it.

ACKNOWLEDGEMENTS

Thanks to my supervisors, Yonatan Shemmer and Eric Olson, for their help at all stages of my PhD. Their comments and suggestions have played a large part in shaping what is at last the finished product, and which have without doubt improved my thesis immeasurably. Thanks also to Simon Kittle, whose perceptive and challenging comments on the first three chapters led to some rewriting and just a little soul-searching.

I would like to extend my gratitude to Saul Smilansky and Manuel Vargas. Their thoughts and advice—offered in email correspondence and in person—were of much assistance when it came to writing about their theories. I owe them a debt also in terms of the influence their theories had on my own thinking about free will and moral responsibility.

This thesis is dedicated to my wife, Yuko, whose presence throughout has made the whole thing worthwhile.

CONTENTS

“Life calls the tune, we dance”

John Galsworthy, 1910

Introduction

This thesis came into being from a desire on my part to achieve three interrelated goals. The first of these was to help rekindle a general interest in a principle that I felt has come to be undeservedly neglected in recent times.[1] The principle at issue is the *Principle of Sufficient Reason* (hereafter the *PSR*), first expressly formulated in the 17th century by Leibniz, and implicitly appealed to on occasion prior to that point. To quote one of Leibniz's formulations, the principle holds that "there can be no fact that is true or existent, or any true proposition, without there being a sufficient reason for its being so and not otherwise."[2] In the present day, it is rare to find the *PSR* being explicitly invoked in the way that figures such as Spinoza and Leibniz did not hesitate to do, yet implicit appeals continue to be made to just such a principle in the arguments of many contemporary philosophers. In other words, even as the principle is looked upon as evidence of rationalism having overreached itself, the intuition to which it gives expression remains alive and well, and as such I believe that its relative neglect is unwarranted.

My second goal in writing this thesis went beyond the first, in that I sought not only to 'raise awareness' of the continuing importance of the *PSR* to our philosophical debates, but also, and unfashionably, I resolved to come down on the side of the principle and affirm its truth. As a consequence of embracing the truth of the *PSR*, I would thereby be committing myself also to affirming the truth of determinism—at least, I would be if I were to embrace the principle in the form that Spinoza and Leibniz present it, which is indeed what I had in mind. I wished to make the case, then, that the *PSR*, as articulated by Leibniz in the preceding paragraph, was true, and that this entailed the truth of determinism (the thesis that the laws of nature, conjoined with any proposition accurately describing the complete state of the world at some instant, entails any other true proposition).

[1] Although it is heartening to note that Pruss's (2011) *The Principle of Sufficient Reason: A Reassessment* was published soon after I began this thesis, providing the first monograph devoted to the principle for at least half a century.
[2] Leibniz (1956a), p. 646.

While the first two goals are metaphysical in their concerns, my third goal shifts the latter part of this thesis into the territory of ethics. The question I had in mind to answer was: what, if any, are the consequences of embracing determinism in terms of our ability to legitimately ascribe one another with the free will required for moral responsibility? More particularly, I wanted to know whether, as incompatiblists argue, the truth of determinism precludes the possibility of legitimately ascribing moral responsibility, or whether instead a compatibilist solution can be found. The short answer, as my thesis reveals, is that I concluded that it is possible to provide a satisfactory compatibilist solution to this problem, although the compatibilist solution I offer is by no means an orthodox one. While I accept that ascriptions of moral responsibility are sometimes justified, I argue that they can only be so on consequentialist grounds, since incompatibilists are right to think that no-one is truly deserving of the praise, blame, reward, or punishment they receive.

The pursuit of these three goals spans the course of seven chapters. Chapter 1 is essentially a work in the history of philosophy, offering a survey of the *PSR* from antiquity to the present day. We begin by looking at figures such as Anaximander and Archimedes, to see how explicability intuitions informed the arguments of the ancient Greeks at a time when the *PSR* had yet to be afforded the status of a principle. Following this, we proceed chronologically to the Middle Ages, then on to the era of Spinoza and Leibniz in the 17th century as Leibniz provides the first formulation of the principle. Finally, we move into the present day, and a time when the *PSR*'s stock appears somewhat diminished. Nonetheless, the principle remains in use by contemporary philosophers, sometimes appearing explicitly (for instance as a premise in certain cosmological arguments) and at other times implicitly (such as in a thought experiment by Parfit that examines personal identity).

Chapters 2 and 3 together provide the case for thinking that the *PSR* is true. Chapter 2 examines a clutch of arguments both for and against the principle in a bid to determine whether it should be accepted or rejected. Arguments in favour of the *PSR* are adduced by Leibniz, Wolff, and Pruss, while the case against is represented by arguments from Hume and van Inwagen, as well as an argument that our knowledge of quantum theory demonstrates that the principle must be false. Although the arguments both for and against are found to fall short of providing either confirmation or refutation of the principle, the case is made that there is evidence

enough for its truth that we should adopt a presumption in its favour. This chapter also makes the case for adopting the sort of robust reading of the *PSR* of which Spinoza and Leibniz would no doubt approve, and argues that any such reading entails a commitment to the truth of determinism. Finally, it is proposed that the *PSR* sceptic's best hope for undermining the principle lies in offering empirical evidence for some demonstrably non-deterministic event.

It is the possibility of establishing the reality of non-deterministic events that motivates Chapter 3, which considers the theory of libertarianism as a potential counterexample to the principle. The subject matter of this chapter—i.e. libertarianism—heralds an imminent change in direction for this thesis as we shift from the metaphysical concerns of the *PSR* to the ethical concerns of theories of free will. Before making this shift, however, the truth of the *PSR* remains in the balance, since if libertarians can provide a convincing case for the indeterminism that their theory posits then we can infer from this the falsity of the principle. To cut a long story short, it is found that libertarianism cannot offer a convincing case for indeterminism, and so the presumption in favour of the *PSR* that was established at the close of Chapter 2 prevails.

A slightly less abridged version of the same story is that the libertarian theories of Ginet and Kane are subjected to two arguments, the first of which does not depend on the truth of the *PSR* while the second one does, and it is found not only that each argument on its own provides sufficient reason for rejecting Ginet's and Kane's particular theories, but that these arguments can be used as ammunition against any and all libertarian theories. In response to any accusations that, in assuming the truth of the *PSR*, the second argument is committing the cardinal philosophical sin of begging the question against the libertarian, the response is that no such transgression takes place: the libertarian's empirical evidence for indeterminism is simply too weak to justify their rejection of the principle, and hence its employment as a premise in an argument against libertarianism is justified. It cannot be said that the deliberations of Chapters 2 and 3 deliver to us the verdict that the *PSR* is incontestably true; but they certainly tip the balance in its favour. Not only that, they also provide us with sufficient reason to ponder the ramifications of the truth of a corollary of the principle, namely determinism.

The fourth chapter marks the beginning of what can be considered the second half of the thesis, as the focus shifts from the metaphysical to the ethical. The question now at stake is what implications the truth of determinism has for our moral responsibility practices, with this chapter focusing on arguments for the sceptical conclusion that moral responsibility and determinism are incompatible. Van Inwagen's famed Consequence Argument is the first to be appraised, and, in spite of the ingenious critiques of Lewis and Slote, the argument is judged to be sound: as the argument's conclusion expresses it, no one has power over the facts of the future, including one's own actions. It is argued that such an admission is equivalent to a repudiation of the claim that we enjoy the free will required for moral responsibility, which in turn suggests that we should abandon our moral responsibility practices so as to reflect this reality.

As if this were not enough, a further argument provides grounds for believing that moral responsibility is impossible irrespective of whether determinism is true. The argument in question is Galen Strawson's aptly named Impossibility of Moral Responsibility Argument, which makes the case that, since we act as we do because of our character, and since we cannot be held ultimately responsible for our character, neither can we be held ultimately responsible for our actions. Once more, and notwithstanding an interesting critique from Hurley, it is concluded that the argument is successful. At the close of Chapter 4, then, there seems to be a strong case for favouring the abandonment of our moral responsibility practices. Another finding from this chapter is that what unites these two anti-compatibilist arguments is an intuition in favour of what has been termed the origination condition, which states that, in order to be morally responsible for an action, the source of that action must be in the agent performing it rather than in something external. While compatibilists can and do deny the need for origination, I argue that the denial of such an intuitive condition is simply not credible: the origination condition must be met if we are to be ultimately morally responsible.

Chapter 5 finds us stuck between a rock and a hard place: the arguments of the previous chapter suggest that moral responsibility and determinism are incompatible; yet simply dispensing with moral responsibility seems a drastic and unappealing course of action to recommend. With traditional compatibilist theories looking indefensible in the light of the arguments of Chapter 4, and hard determinist theories

looking impracticable and undesirable in equal measure, the next two chapters are devoted to finding a way of steering a course between these compatibilist and hard determinist poles. Doing this involves examining what may be called 'mixed views' on moral responsibility, a name that derives from the fact that each view combines elements of both compatibilist and incompatibilist thought. The hope is to find a theory that offers both a satisfying response to the origination problem (unlike traditional compatibilist theories), and at least some justification for our moral responsibility practices (unlike hard determinist theories).

With these goals in mind, Vargas's revisionism is the first theory under discussion. It is argued that Vargas is right to make the case that we are folk libertarians, and that we are mistaken in this regard; but his prescription is flawed, since this turns out to be little more than the recommendation that we adopt the kind of compatibilism that involves denying the necessity of meeting the origination condition. However, Vargas also offers the helpful proposal that our responsibility norms serve an essentially utilitarian function, which is to improve us morally over time. While Vargas himself stops short of adopting a consequentialist justification of moral responsibility, I argue that this approach allows us to acknowledge the necessity of origination for ultimate responsibility without this obliging us to abandon our moral responsibility norms.

The other mixed view examined in Chapter 5 is Double's free will subjectivism, according to which judgements concerning moral freedom cannot be objectively true. While free will subjectivism certainly offers a simple justification for our moral responsibility judgements—that is, in the absence of any objective measure of moral freedom we are at liberty to pick and choose when to ascribe it—this approach turns out to be too permissive to be satisfactory. For one thing, Double's theory fails to provide an adequate response to the origination problem, effectively saying that it is up to us whether we choose to make a connection between it and moral responsibility.

The final mixed view under consideration is Smilansky's, to whose theory most of Chapter 6 is devoted. Smilansky makes two separate proposals, the first of which is that we must adopt a 'fundamental dualism' with regards to the theories of compatibilism and incompatibilism, while the second is that we recognise and

embrace what he terms ‘illusionism’ with respect to free will. Dealing first with fundamental dualism, this is the notion that neither compatibilism nor incompatibilism is adequate on its own, although both have aspects that are indispensable. Therefore, he argues, we must retain those parts of each that cannot be discarded, and abandon the common assumption that either compatibilism or incompatibilism must be correct in its entirety.

I contend that, although Smilansky is right that we need a fundamental dualism, we should not adopt one in the form that Smilansky himself suggests. Smilansky’s approach is too equivocal about the success or otherwise of arguments that view origination as a necessary condition for ultimate responsibility, and so he adopts a position whereby lack of origination is sometimes a crucial factor while at other times it is not. There is no clear case for thinking that origination is sometimes not necessary, and so Smilansky’s fundamental dualism falters. However, I propose that the idea that motivates Smilansky’s fundamental dualism—that neither compatibilism nor incompatibilism is adequate on its own—is worthy of consideration, and that we should maintain a dualism while changing the details. Briefly put, I argue that compatibilist judgements of moral culpability provide a consequentialist justification for our moral responsibility practices, while incompatibilist judgements show that no-one is ever ultimately deserving of being held morally responsible.

As for Smilanksy’s proposal of illusionism, this combines the claim that the majority of people have illusory beliefs concerning free will with the further assertion that this situation is for the best. Smilansky’s first claim, like Vargas’s, is simply that folk libertarianism is prevalent and that they are mistaken on this point. This much can be accepted. The further assertion that this is for the best, however, is less secure, as the supposed threat posed by the unmasking of this illusion is somewhat overstated. Still, the possibility that illusion might play a useful part in a theory of moral responsibility is conceded, if knowing the truth should prove to be too much to bear. We end the chapter with a brief consideration of Smart’s consequentialist justification of moral responsibility practices, whose theory allows us to affirm the necessity of the origination condition for true desert without abandoning hope that we can continue to hold one another morally responsible.

Chapter 7 concludes the thesis, and makes the positive case for adopting a consequentialist justification of our moral responsibility norms in the face of the most compelling reasons for not doing so. This chapter also summarises how and to what extent aspects of the various mixed views surveyed in Chapters 5 and 6 might be useful in developing the present account of moral responsibility, an account that avoids going down either traditional compatibilist or incompatibilist routes.

I finish this introduction by offering something of a disclaimer. This is to acknowledge that the connection between what I have referred to as the first and second halves of my thesis (Chapters 1-3 and Chapters 4-7 respectively) is not as seamless as it might have been. A more pleasing thesis structure would certainly have been achieved if Chapter 4 had delivered the result that moral responsibility is impossible *because of* the truth of determinism. While van Inwagen's argument certainly makes this connection between determinism and moral responsibility, whereby the truth of the former precludes the possibility of the latter, Galen Strawson's argument has it that moral responsibility is impossible *regardless of* the truth or falsity of determinism. Given that I accept the soundness of Galen Strawson's argument as well as van Inwagen's, one might be forgiven for asking: what need is there to attempt to establish the truth of determinism via a defence of the *PSR* before outlining a theory of moral responsibility?

In response, I think establishing the truth of determinism before outlining a theory of moral responsibility finds a measure of justification in the fact that there will be some readers who will deem van Inwagen's argument persuasive but not Galen Strawson's. For such readers, establishing the truth of determinism will be necessary in order to convince them of the need to pursue a mixed view theory of moral responsibility. Equally, of course, some may find Galen Strawson's argument the more persuasive of the two, especially those who have reservations about the strength of the case for the *PSR* and determinism. In accepting Galen Strawson's argument, these readers should still be able to agree with my account of how to proceed in developing a viable theory of moral responsibility. Each argument is useful, in other words, where misgivings exist regarding the other one.

1

A survey of the *Principle of Sufficient Reason*

It is a truth frequently affirmed that the *Principle of Sufficient Reason* (hereafter *PSR*) is as old as philosophy itself. Indeed, ever since the time of the Ancient Greeks the intuition has been invoked that, given some particular fact or event, there must necessarily be a reason for its obtaining or occurring. Still more often, the intuition that facts and events are in principle explicable has simply been implied although left unstated. This intuition which the *PSR* seeks to formalise persists through the Middle Ages, during which era such great philosopher-theologians as Abelard and Aquinas appeal to it, respectively, in arguments for 'optimality' (the idea that this is the best of all possible worlds), and in proofs for the existence of God.

However, it is not until Leibniz in the enlightenment era that the intuition that facts and events should be explicable becomes promoted to the status of a principle. It was Leibniz who duly coined the term '*Principle of Sufficient Reason*,' which was judged by the great polymath to be one of two great principles upon which all truths rest (the other being the *Principle of Contradiction*, according to which a truth is necessary just in case its negation is a contradiction). Additionally, Leibniz can be credited with having devised the first known argument for the principle, which seeks to demonstrate that whatever is—and whatever is true—must have a sufficient reason. At last, the intuition that philosophers have for centuries appealed to finds itself articulated clearly and explicitly by Leibniz, and, what is more, an argument is adduced as evidence of its proof.

Sadly, as is so often the case in philosophy, matters did not remain simple and settled for long. A generation later, Hume led the charge against the *PSR* and the intuitions undergirding it by claiming that there is no contradiction in the idea that a thing might lack a cause or reason for its existence, and hence that the common intuition

that this is not possible is false. Since then, as Della Rocca remarks despondently, "it sometimes seems as though […] a great deal of the best efforts of the best philosophers have been devoted to a direct frontal assault on the PSR."[1]

Nevertheless, despite the common presupposition among contemporary philosophers that the *PSR* is false, the situation is far from a consensus. In fact, philosophers continue to construct arguments that appeal, whether implicitly or explicitly, to Leibniz's great principle. Parfit's famous discussion of personal identity and fission involves an implicit appeal to the *PSR*, while contemporary philosophers of religion such as Reichenbach appeal explicitly to the *PSR* when arguing for the existence of God.

In this first chapter, then, let us perform a brief survey of the *PSR*, looking in particular at how the principle has been applied. While there is not space to present an exhaustive history of the *PSR*'s application, I do hope to achieve the more modest aim of providing an understanding of many of the key uses of the *PSR*, from the time of the ancient Greeks up until the present day.

1.1 Ancient Greeks

Anaximander is usually credited with being the first person to employ the *PSR*. Aristotle's *De Caelo* records Anaximander as arguing that the earth remains stationary in space since it is indifferent between motions in any particular direction.[2] It is indifferent since it is "equably related to the extremes," and so there can be no reason for it to favour movement in one direction rather than another. So, in the absence of a reason why it should move in one direction rather than in any other, Anaximander concludes that it "necessarily remains at rest." More generally, Anaximander seems to be assuming that *motion in the absence of a reason is impossible*.[3]

[1] Della Rocca (2010), p. 2.
[2] Aristotle (1939), b12 295b10–16.
[3] Aristotle does not uncritically report Anaximander's argument – on the contrary, he considers Anaximander's reasoning to be "ingenious, but not true." However, it is not that Aristotle denies the *PSR*-type assumption that motion in the absence of a reason is impossible: he simply identifies a *different* reason from Anaximander for the earth's being stationary, which is that "motion to the centre

This same assumption—that motion in the absence of a reason is impossible—is employed to different effect by Archimedes. This became Leibniz's favoured example of a historical application of the *PSR*, one that he would often cite in correspondences and other philosophical writings.[4] Leibniz reports of Archimedes: "[He] takes it for granted that if there is a balance in which everything is alike on both sides, and if equal weights are hung on the two ends of that balance, the whole will be at rest. That is because no reason can be given why one side should weigh down rather than the other."[5] Archimedes thus appeals to the thought that, in the absence of a reason why one side should weigh down rather than the other, the whole will be at rest.

Parmenides' use of the *PSR* is more ambitious than that of his contemporaries.[6] While Anaximander and Archimedes were concerned to derive the consequences of the claim that *motion* in the absence of a reason is impossible, Parmenides concerns himself with the altogether bolder claim that *existence* in the absence of a reason is impossible. Since there could be no reason (Parmenides argues) for something to come from nothing, we can be assured that nothing does—or even could—come from nothing. This is a statement of the *ex nihilo nihil* principle, which simply states that no entity comes into existence out of nothing.

So what argument is offered for this assertion that there could be no reason for something to come from nothing? Of the putative thing that comes from nothing, Parmenides asks: "[W]hat need would have driven it later rather than earlier, beginning from the nothing, to grow?"[7] The rhetorical point Parmenides seems to be making is that, if (*per impossible*) something were to come to exist from nothing, we would require an explanation as to why it came to exist when it did, rather than, say,

---

is peculiar to earth" (as opposed to, say, fire, whose nature is to move to the extremes). From our vantage point, of course, we can see that *both* Anaximander's *and* Aristotle's explanations are faulty, since the supposition that the earth is stationary in space is false.

[4] Two reasons can be adduced for Leibniz favouring this Archimedean usage of the *PSR* above other prospective candidates: first, it seems like a particularly uncontroversial application of the *PSR*; second, Archimedes' application involves no faulty assumptions. Regarding the first point, Parmenides' application of the *PSR*, for example, is manifestly more controversial than Archimedes: whereas Archimedes assumes that *motion* – or, more specifically, the movement of a balance—requires an explanation, Parmenides is committed to the *prima facie* stronger claim that the *existence* of things must be explicable. As to the second point, we can note that Anaximander, ingenious though his argument may be, is of course incorrect in his assumption that the Earth is stationary in space.

[5] Leibniz (1989), p. 31.

[6] In actual fact, a good 350 years separate Anaximander and Archimedes; but since this gulf in time is far less apparent from our vantage point I feel justified in describing them as "contemporaries"!

[7] Parmenides (1986), Fr. 8, 9-10.

twenty minutes earlier. But of course no such explanation could be given, since all empty times are homogenous.

Note that, although Parmenides' application of what we might call an "explicability argument" is bolder than either Anaximander's or Archimedes', it is clear that their arguments possess a similar form and give expression to a similar intuition: Anaximander makes the claim that the earth is indifferent to motion in any direction—and so lacking a reason to move, it naturally remains motionless; Archimedes describes a balance in which two sides are identical in every respect (they have the same object, of the same mass, placed at the same distance from the fulcrum)—and so lacking a reason to tip one way or the other, it naturally rests in equipoise; Parmenides makes the observation that empty times are identical in every respect—and so lacking a reason to come into existence at one empty time as opposed to another, nothing *can* come into existence from an empty time. In short, for everything from the mundane to the remarkable, reasons are required. The intuition is: explicability rules.

This intuition also finds expression in Plato and Aristotle. Plato offers us what can be considered a proto-*PSR* when he states: "It is necessary for everything that happens to happen through a cause; for how could it happen without this?"[8] Similarly, Aristotle provides the following reflection: "We think we understand everything perfectly when we think we know the cause whereby the thing exists, namely that it is the cause of that thing, and that this could not possibly be otherwise."[9] Of course, both of these statements speak of causes as opposed to reasons; but as I shall argue later in this chapter, the term 'cause' can be considered synonymous with the term 'reason', at least insofar as concerns the reasons for spatiotemporal events.[10]

[8] Plato (1971), 28a, 4–5.
[9] Aristotle (1994), I, 2.
[10] In fact, Barnes's translation of Aristotle (1994: p. 2) renders 'cause' as 'explanation,' which seems unobjectionable, since in the quoted passage Aristotle is concerned with our *understanding* of things.

1.2 Middle Ages

Use of the *PSR* continues into the medieval period. Abelard, the French scholastic philosopher, theologian, and logician, is responsible for conceiving of the notion that we inhabit the best of all possible worlds (a full 500 years before Leibniz, whose name is most commonly associated with the idea, proposed the same idea).[11] His argument is that we must inhabit the best of all possible worlds because God, naturally wanting the best for his creation, would have no reason to *fail* to create the best of all possible worlds. In the absence of any reason *not* to create the best of all possible worlds, the best of all possible worlds is therefore created.[12]

Unfortunately for Abelard, his rational bent had the effect of antagonising church council members of a more mystical leaning, who objected (among other things) to his predilection for quoting gentile philosophers.[13] Abelard's defence of this practice, following a church council at Sens in 1141 at which Bernard of Clairvaux charged him with heresy, came in a screed arguing for the legitimacy of these sources. The following passage is particularly noteworthy, as it illustrates both his apparent commitment to the *PSR* and to the doctrine of 'optimality' (or, as Leibniz would later refer to it, the *Principle of the Best*):

> [The Greeks] postulate 'the highest good' (*summum bonum*). This is God, the beginning of all things, the origin and efficient cause of everything. Thus they insist that from love of Him all goodness, just as everything else, must proceed. We, in a similar way, call God alpha and omega, the beginning and the end, from Whom are all things, and on Whose account all things are in being. Plato calls Him the highest and ineffable Creator of all natural things, Who is able to do everything and from Whom all evil is removed. He has formed all good, each thing according to its nature or as its order and harmony required. [14]

---

[11] My phrasing here (stating that Abelard is "responsible for" optimality theory) is intentionally ambiguous. Many would argue that it is to Abelard's discredit that he argued that this is the best of all possible worlds, although it is an idea that continues to have some currency. For modern defences of this theory, see Leslie (2001); and Rescher (2000).

[12] Abelard's theology is reflected not only in his philosophical prose, but also in some of his poetry. The following stanza from the poem *Morning Hymn* is further illustration of his belief in optimality and the rule of reason: "Perfect in every part / Thy perfect world began; / In every part endures / In reason's faultless plan." (quoted in McCallum (1976) p. 114)

[13] One of those 'other things' to which his fellow philosophers of religion took exception was in fact the doctrine of optimality, here expounded. The mainstream view among medieval philosophers was that God enjoys freedom of indifference with respect to his creation. As such, God is not generally seen as being compelled by his nature to create the 'best of all possible worlds.'

[14] Abelard, quoted in McCallum (1976), p. 61.

There are two key points to note from this passage. First, there is Abelard's implicit commitment to the *PSR* as revealed in the assertion that all things "must proceed" from God: that is, God is declared to be the origin and efficient cause of everything - and hence also the sufficient reason for the existence of everything. Second, we can observe Abelard's endorsement of the doctrine of optimality, according to which God, being the *summum bonum* or highest good, has "formed all good, each thing according to its nature or as its order and harmony required." In other words, the sufficient reason for God creating what he in fact creates is that, in keeping with his character as the highest good, it is the best creation possible. We will see that Abelard's claim—that the truth of the *PSR,* together with the existence of a perfectly good God entails optimality—is reiterated by Leibniz in the 17th century, and we shall see also how Leibniz struggles to reconcile this claim with the notion of contingency. Before this, however, there is another medieval philosopher who makes notable use of the *PSR*, and whose influence as a forebear to Leibniz is also apparent. That philosopher is Thomas Aquinas, and his arguments for the existence of God—which all fall under the general rubric of 'Cosmological Arguments'—will now be briefly examined. A more recent statement of the cosmological argument will also be examined presently, when we consider contemporary applications of the *PSR*.

Aquinas presents 'Five Ways' of demonstrating the existence of God. The general form of each argument is the same: first, some fact about the world is cited; second, it is claimed that this fact requires an explanation; third, the conclusion is that the only sufficient explanation for the fact in question is God. So, the *PSR* appears as a premise in each of the Five Ways, in the form of a demand for an explanation for each particular fact. The Five Ways are as follows: the first way concerns change, arguing that there must be an Unmoved Mover that originates all change; the second way argues that there must be a first cause to explain the existence of all other causes; the third way argues that because contingent beings exist (e.g. humans) there must be a necessary being who explains the existence of contingent beings; the fourth way notes the existence of degrees of excellence, and so posits a perfect being as the explanation as to the source of these excellences; lastly, the fifth way asserts

that the harmony of nature requires explanation, and that only a divine designer could sufficiently explain such a fact.[15]

These cosmological arguments have come under attack in a great variety of forms since being propounded by Aquinas. However, for many critics, it is the truth or falsity of the *PSR* that is the key to the arguments' success or failure. Perhaps the most durable of the Five Ways is the third, which offers an argument from contingency—Leibniz and Clarke both develop versions of this, and the debate continues today as to the soundness of some such formulation of the cosmological argument. [16]

As can be seen, none of the Five Ways in fact invokes the *PSR* in its full generality: that is, rather than baldly stating the general claim that every fact or event has a sufficient reason, Aquinas instead appeals to our intuition that, for *this particular fact or event*, there should be a complete explanation. Still, it seems likely that Aquinas himself would assent to the full-blown *PSR*, given the wide array of facts and events that he seems to assume do require an explanation. By contrast, Spinoza and Leibniz, enlightenment-era philosophers of a distinctly rationalist cast of mind, affirm the *PSR* explicitly and unashamedly, and afford the principle a central place in their respective philosophies. We will now examine how they applied the *PSR*, and the conclusions they derive from the principle.

1.3 Spinoza and Leibniz

Spinoza, while he cannot be credited with coining the term '*Principle of Sufficient Reason*,' certainly makes extensive and explicit use of the principle. Spinoza phrases the *PSR* as follows: "For every thing a cause or reason must be assigned either for its existence or for its non-existence."[17] The most striking feature of Spinoza's application of the *PSR* is that he takes it to entail necessitarianism. This is the

[15] The Five Ways appears in both Aquinas's *Summa Theologica* (I, q.2) and his *Summa Contra Gentiles* (I, 13). For a concise summary of Aquinas's cosmological arguments, see Pojman's (2003) anthology, pp. 3-5.
[16] Leibniz (1991) develops the cosmological argument in *The Ultimate Origination of Things* (G VII 302–3; L 486–8); and in *Monadology* (§37). See Rowe (1975: pp. 60-167) for a detailed account and critique of Clarke's exposition of this argument.
[17] Spinoza (1992), p. 37.

metaphysical thesis that denies all mere possibility, and says that there is only one way that the world could be (i.e. how it in fact is).

According to Spinoza, one class of objects (A) necessarily do not exist, the reason being that their very nature involves a contradiction. Examples of such things include square circles, or female bachelors. Next, we are told that another class of objects (B) exist of necessity on account of their nature, and these objects Spinoza terms 'substances'. In fact, there is only one substance, and that is God. To say that God exists on account of his nature is to say that God's very nature involves existence (since substances are self-caused, their essence necessarily involves existence).[18] Further, there is nothing external to God that could prevent him existing, because "a substance of another nature would have nothing in common with God, and so could neither posit nor annul his existence."[19]

The last two classes of objects are those for which the reason for their existence or non-existence follows from "the order of universal, corporeal Nature."[20] That is, these objects are ones whose nature—and hence their existence—does not involve a contradiction; but neither does their non-existence involve a contradiction (unlike in the case of God). For such objects, it either "necessarily follows" that they exist, or else their present existence is "impossible" if there be some reason or cause preventing their existence.[21] In short, their existence or non-existence is fully determined by external factors. These two categories, then, comprise (C) every single existing object in the world—people, houses, animals, trees, planets etc.—or, to use Spinoza's terminology all the various 'modes' of God, and (D) every non-existent but not by nature contradictory object in the world—unicorns perhaps, or the Loch Ness monster, or spaghetti trees, and so forth.

In summary, Spinoza uses the *PSR* to posit four categories of object, thereby creating the basis for his ontology. The four categories are:

(A) Objects that are logically impossible (for example, square circles), whose nature involves a contradiction.

[18] *Ibid.*, Pr. 7 Proof, p. 34.
[19] *Ibid.*, Pr. 11 Second Proof, p. 37.
[20] *Ibid.*, Pr. 11 Second Proof, p. 37.
[21] *Ibid.*, Pr. 11, Second Proof, p. 37.

(B) Objects that are logically necessary (that is, God), whose non-existence would involve a contradiction.

(C) Objects that are logically possible (for example people and books), and whose existence necessarily follows from the order of universal corporeal nature.

(D) Objects that are logically possible (again, people, books, and the like), but whose existence necessarily does not follow from the order of universal corporeal nature.[22]

For Spinoza, we can see that it is a consequence of the *PSR* that everything that exists does so of necessity; and conversely, everything that does not exist, does not exist of necessity. Objects in categories (B) and (C) exist of necessity; objects in categories (A) and (D) necessarily do not exist.

As Spinoza had done just a few decades before him, Leibniz also affords the *PSR* a central place in his philosophy—and as Spinoza had, Leibniz also uses the principle to argue for the existence of God.[23] But Leibniz also makes novel use of the *PSR*. For example, he employs it to derive the principle of the *Identity of Indiscernibles*, a principle which precludes the possibility of there being two or more things that are exactly alike. The reason Leibniz adduces for the impossibility of there being two indistinguishable things is that, if (*per impossible*) two indistinguishable things were to exist, then God should have no reason to put them in different relations to the rest of the world from each other. Since nothing could explain their different relations to

[22] The claim that the objects of categories (C) and (D) either exist or fail to exist necessarily despite not being *logically* necessary begs the question (in the non-philosophical sense of that phrase): in what sense *are* they presumed to be necessary? The answer, I believe, is that the objects of categories (C) and (D) should be considered *metaphysically* necessary. This option is advocated by Pruss, a present day *PSR* proponent. Precisely how the notion of the metaphysical necessity of some objects is to be elucidated, however, beyond simply stating that they follow from the order of universal corporeal nature, is not so clear.

[23] Spinoza's basic idea is as follows: if God did not exist, then there would be some cause or reason for his non-existence; no cause or reason, either internal or external to God, could prevent him from existing; therefore, God necessarily exists. For an excellent exposition of all four of Spinoza's arguments for the existence of God (three of which employ the *PSR*) see Lin's (2007) article "Spinoza's Arguments for the Existence of God." Leibniz's line of thought is somewhat different, and is more in keeping with traditional cosmological arguments. We can summarise his argument thus: as the *PSR* insists, contingent things require an explanation; the explanation of the series of contingent things cannot itself be part of this series, since then it would be self-explanatory (and contingent things by definition cannot be self-explanatory); so, the explanation must be something necessary and, given that any necessary being is God, it must be God. See fn. 16 for citations of Leibniz's cosmological argument.

the world, it simply cannot be the case that two indiscernible things exist.[24] A further novel use of the *PSR*, and one that is in fact a consequence of him adopting the principle of the *Identity of Indiscernibles*, is that Leibniz infers from it that space is relational. According to Leibniz, space and time cannot themselves be substances, but must rather be merely a system of relations that obtain between bodies. The reason for this is that, if space were to be absolute, then it would be the case that different points in space would be exactly alike: that is, they would be indiscernible from one another.[25] Once again, God could have no reason to treat one point in space differently from another, indiscernible, point in space. In such a situation, God would have to make an arbitrary decision about how to order the universe, and this (as per the *PSR*) cannot be countenanced.[26]

How exactly did Leibniz express the *PSR*? Throughout his philosophical career Leibniz offers many statements of the principle, a characteristic one being:

"[T]here can be no fact that is true or existent, or any true proposition, without there being a sufficient reason for its being so and not otherwise, although we cannot know these reasons in most cases."[27]

This statement of the *PSR* is noteworthy for a couple of reasons. First, the scope of the *PSR* construed here is extremely broad, since not only does Leibniz state that there is a sufficient reason for the existence of things (as does Spinoza), but he declares further that there is a sufficient reason for any true proposition (a claim not explicitly made by Spinoza). In the case of propositions that refer to "necessary"[28] truths (such as mathematical and metaphysical truths), the sufficient reason for their truth, Leibniz claims, is that their negation is a contradiction.[29] The second noteworthy point about Leibniz's formulation here is the last nine words, which serve to qualify the sentence. These make it clear that the *PSR* is not a claim that we can *know* the reason for any given event or truth, simply that there *is* a reason—that is, it is an ontological claim rather than an epistemic one.[30] The *PSR* as construed by

[24] Leibniz (1989), p. 42.
[25] Leibniz (1956b), 3.5.
[26] For an aggravatingly obtuse discussion of Leibniz's relational theory of space, see Belot (2001), pp. 62-70.
[27] Leibniz (1956a), p. 646.
[28] I use this term advisedly, since Spinoza of course maintains that *all* truths are necessary, those which refer to ostensibly contingent facts every bit as much as mathematical and metaphysical truths.
[29] See, for example, *Monadology* §36.
[30] This is not how the *PSR* has always been understood, which is why it is important to make it clear that this it is the Leibnizian definition which is of relevance to this thesis. An example of an epistemic

Leibniz, then, does not entail that we can (even if only in principle) always discover the sufficient reason for some truth or other: however, it does entail that, if some truth is known then there must be a sufficient reason for its being known.

While Leibniz of course affirms the *PSR* and derives some interesting conclusions from it as detailed, he denies that it entails necessitarianism. Leibniz argues, as Abelard had before him, that our world is but one among many other possible worlds, chosen by God on the basis of the much-ridiculed *Principle of the Best*.[31] This *Principle of the Best* governs all "contingent" truths, as God supposedly created this world for the reason that it was the best among all possible worlds. This claim enables Leibniz (or so he believes) to avoid necessitarianism, since other, non-actual, worlds were potential candidates for actualisation. Leibniz elaborates as follows: propositions, he claims, are either true by absolute necessity (in which case their negation is a contradiction), "or by a kind of certainty which depends upon the supposed decree of a free substance in contingent matters, a decree, however, which is never entirely arbitrary and free from foundation, but for which some reason can always be given. This reason, however, *merely inclines and does not truly necessitate*."[32] (My italics)

In response to this attempt to avoid necessitarianism, we should note that if God chose this world according to the *Principle of the Best*, then the choice was surely a necessary one, since the fact that this world is the best among possible worlds would necessitate God's choice. A further point to note is that Leibniz's assertion that sufficient reasons often incline without necessitating runs counter to a definition that he offers for the term 'sufficient reason' in the context of an argument for the *PSR*.

---

reading of the *PSR* is Wolff, who, according to Kant, "defines reason (or ground) as that from which it is possible to understand why something is rather than is not." (Longueness (2001) p. 69) A further example of an epistemic reading of the *PSR* comes from Kiesewetter: "Logical ground or reason (reason of knowledge) is not to be confused with the general ground or reason (cause). The principle of sufficient reason belongs to logic, the principle of causality belongs to metaphysics. The former is the fundamental principle of thought, the latter that of experience. Cause concerns actual things, logical reason or grounds concerns only representation." (cf. Schopenhauer (1974), p. 30)

[31] Most famously ridiculed by Voltaire of course, whose 1759 novel *Candide* satirises Leibnizian optimism by showing Dr. Pangloss, a teacher of Leibniz's doctrine, blithely proclaiming the platitude that "all is for the best in the best of all possible worlds" in the face of mounting evidence to the contrary. Hilarity ensues.

[32] Leibniz (1956a), p. 226.

Here, it is defined as follows: "A sufficient reason is that which is such that if it is posited the thing is."[33]

In summary, according to Leibniz himself, sufficient reasons never merely incline but rather by definition entail their *explananda*. Furthermore, if the *Principle of the Best* does indeed govern all supposedly contingent phenomena as Leibniz claims, then it seems to be impossible to avoid the very conclusion that Leibniz is so anxious to avoid - namely that everything happens of necessity. Further discussion of this issue will be offered when it comes to examining van Inwagen's modal argument against the *PSR*; but let us leave the discussion of necessitarianism for now and move on to investigate the use of the *PSR* in contemporary philosophical arguments.

1.4 Modern Day

The *PSR* has fallen out of favour somewhat since the era of Spinoza and Leibniz, a situation alluded to earlier when quoting Della Rocca's observation that the *PSR* has faced a "direct frontal assault" in both recent and not-so-recent years. Some of these arguments against the *PSR* will be appraised shortly, but before examining these we would do well to note that, notwithstanding this "assault," the *PSR* is still very much in use. This usage is either explicit or else it is implicit, and an example of each will now be sketched.

For a well-known argument in which commitment to the *PSR* (or to some sort of explicability principle at least) is in evidence, we need look no further Parfit's writing on personal identity.[34] Parfit presents us with a thought experiment, which he terms "My Division." We are asked to imagine that Parfit is an identical triplet, to whom the following unbelievably dire situation occurs:

> My body is fatally injured, as are the brains of my two brothers. My brain is divided, and each half is successfully transplanted into the body of one of my brothers. Each of the resulting people believes that he is me, seems to remember living my life, has my character, and is in every other way psychologically continuous with me.[35]

---

[33] Adams (1994), p. 68.
[34] Parfit (1984), pp. 239-259.
[35] *Ibid.*, pp. 254-5.

In response to the obvious thought that such a scenario is impossible (after all, no such brain transplant has ever been—or likely could ever be—performed) Parfit justifiably responds that this impossibility is "merely technical."[36] That is, it is not logically impossible that such a procedure should happen, and in fact "[t]he one feature of the case which in fact might be considered *deeply* impossible—the division of a person's consciousness into two separate streams—is the feature that has actually happened."[37]

So, if we accept this hypothetical scenario as Parfit urges we must, we are faced with the question of what has become of Parfit. The options, Parfit suggests, are as follows:

(1) Parfit does not survive.
(2) Parfit survives as one of the two people.
(3) Parfit survives as the other person.
(4) Parfit survives as both.

It is options (2) and (3) that interest us here, as it is to these options that the *PSR* is employed in order to conclude that neither one provides a satisfactory answer. In Parfit's own words:

> Consider the next two possibilities [i.e. (2) and (3)]. Perhaps one success is the maximum score. Perhaps I shall be one of the two resulting people. The objection here is that, in this case, each half of my brain is exactly similar, and so, to start with is each resulting person. Given these facts, how can I survive as only one of the two people? What can make me one of them rather than the other?[38]

Parfit seems to be asking: what could possibly count as a reason for claiming that he should survive as one of these two people rather than as the other? Since each half of his brain is exactly similar, there cannot be any reason to plump for (2) as opposed to (3), or *vice versa*. And, in the absence of any reason, it cannot be that Parfit is one of them but not the other. Such an admission of a brute fact would be unacceptable.[39]

---

[36] *Ibid.*, p. 255.
[37] *Ibid.*, p. 255. I am taking Parfit at his word when he says that the division of a person's consciousness into two separate streams has actually happened, since I am unaware of the case myself.
[38] *Ibid.*, p. 256.
[39] It should be noted however that, shortly after presenting his argument, Parfit does consider the possibility that he might become one or another of the two people at random: "In my example, there would be no reason why the particular ego that I am should wake up as one of the two resulting

We move now from an example of the implicit application of the *PSR* to its explicit use, in the form of modern variants on the cosmological argument. It is Aquinas's third way that has proved most resilient, and which receives a contemporary formulation from Reichenbach, who states the argument as follows:[40]

(1) A contingent being exists.
    a. This contingent being is caused either (i) by itself, or (ii) by another.
    b. If it were caused by itself, it would have to precede itself in existence, which is impossible.
(2) Therefore this contingent being (ii) is caused by another, i.e. it depends on something else for its existence.
(3) That which causes (provides the sufficient reason for) the existence of any contingent being must be either (iii) another contingent being, or (iv) a non-contingent (necessary) being.
    c. If (iii), then this contingent cause must itself be caused by another, and so on to infinity
(4) Therefore that which causes (provides the sufficient reason for) the existence of any contingent being must be either (v) an infinite series of contingent beings, or (iv) a necessary being.
(5) An infinite series of contingent beings (v) is incapable of yielding a sufficient reason for the existence of any being.
    Therefore a necessary being (iv) exists. [41]

The *PSR* is central to this argument since, as premise (2) states, a contingent being must be caused by something else, which is to say something else provides the *sufficient reason for* the contingent being's existence. For present purposes, this argument is of interest on account of the debate that it has engendered not only as to

---

people. But this might just happen, in a random way, as is claimed for fundamental particles." (pp. 258-9) It thus seems that Parfit takes consideration of the *PSR* to count as reasonable grounds for rejecting the claim that he would survive in the body of one or the other of his identical triplet brothers; but it is not beyond the bounds of possibility that he should survive in one or the other of his brothers *for no reason whatsoever*. Explanations are to be preferred, then, but cannot for all that be presumed.

[40] Although I have chosen to present Reichenbach's formulation of the cosmological argument, his is by no means the only recent formulation to rely on the *PSR*. Others include: Gale and Pruss (1999); Leftow (1988), and; Meyer (1987). For recent formulations that avoid the *PSR*, see White (1979), and; Katz (1997).

[41] Reichenbach (1972), pp. 19-20.

whether the *PSR* is true, but also as to what *kind* of a truth the *PSR* would be were it to be true.

To examine this latter point first, Clarke, an earlier proponent of this variety of cosmological argument, appears to believe that the *PSR* is a necessary truth. He declares:

> 'Tis easy to conceive, that we may indeed be utterly ignorant of the reasons or grounds, or causes of many things. But, that anything is; and that there is a real reason in nature why it is, rather than not; these two are as necessarily and essentially connected, as any two correlates whatever.[42]

But Clarke's contention no longer holds so much sway, since it is plausible that only analytic truths are also necessary truths, and the *PSR* does not appear to be analytic.[43] Taylor suggests a different account of the status of the *PSR*, arguing that its truth is not necessary but is, nevertheless, a fact that we inevitably presuppose in our thinking. In Taylor's words:

> One can deny that [the *PSR*] is true, without embarrassment or fear of refutation, but one is apt to find that what he is denying is not really what the principle asserts. We shall, then, treat it here as a datum—not something that is provably true, but as something which all men, whether they ever reflect upon it or not, seem more or less to presuppose.[44]

Taylor's characterisation of the *PSR*—that it is a truth that is presupposed in our thinking—leads us onto discussion of the former point, which was the issue of whether the *PSR* is in fact true. First, Taylor's claim that we do presuppose the *PSR* in our thinking is challenged by Rowe, who contends that it is in fact perfectly natural to doubt its truth. Rowe highlights the example of the scientist whom he thinks might very well not presuppose that everything has an explanation in terms of a cause (this possibility will be explored in more depth when we examine the current state of quantum theory).[45] Second, Russell also challenges the assertion that the *PSR* is a presupposition of thought, as he counters that we need not suppose that there is any explanation for the universe as a whole, even if it turns out that there are always explanations within it: in his words, the universe is "just there, and that's

---

[42] Clarke (1998), p. 490.
[43] Rowe (1975: p. 83) expresses this intuition as follows: "The idea of an event, of something happening—a leaf falling, a chair collapsing—does not seem to contain the idea of something *causing* that event. If this is so, then PSR is not analytically true."
[44] Taylor (1963), pp. 86-7.
[45] Rowe (1975), p. 88.

all."[46] Third, Rowe argues that, even if it is conceded that the *PSR* is a presupposition of reason, this is no guarantee of its truth: "The fact, if it is a fact, that all of us presuppose that whatever exists has an explanation of its existence does not imply that nothing exists without a reason for its existence. Nature is not bound to satisfy our presuppositions."[47]

As stated earlier, this survey is not intended as an exhaustive account of the history of the *PSR*'s application, although it is hoped that it will provide an indication of many of its most important uses. In any case, now that the history of the principle has been sketched, we can move on in the following chapter to examining the arguments both for and against it in more detail.

[46] Russell, in Hick (1964), p. 175. This assertion can be traced back to Hume, who argued in a similar vein that the universe as a whole required no explanation over and above its constituent parts, and has been charged (rightly or wrongly) with committing the fallacy of composition (the assumption that explanation of the parts of a thing is sufficient as an explanation of the whole).
[47] Rowe (1975), p. 93.

2

Arguments For and Against the *PSR*

While the job of the previous chapter was to offer a brief historical survey of the *PSR* in order to understand its uses, the present chapter aims to answer the following crucial question: is the *PSR* true? To that end, a total of six arguments will be considered, three of which conclude that the *PSR* is true, while the other three conclude the opposite. Arguments in favour of the *PSR* will be considered first, beginning with the first known formal argument for the principle, articulated by Leibniz. Arguments against include Hume's influential attack on the *PSR*, and an argument that causeless quantum phenomena provide evidence for the falsity of the principle.

The hope is that, by careful examination of the arguments, we might come to know whether we should embrace the *PSR* or else reject it. Not only that, but we should have a clearer idea of what else is entailed by the rejection or acceptance of the *PSR*.

2.1 Arguments For the *PSR*

In his *Fourfold Root of the Principle of Sufficient Reason*, Schopenhauer remarks:

> [T]o seek a proof for the principle of sufficient reason in particular is especially absurd and is evidence of a want of reflection [since] whoever requires a proof for this principle, i.e., the demonstration of a ground or reason, already assumes thereby that it is true; in fact he bases his demand on this very assumption. He therefore finds himself involved in that circle of demanding a proof for the right to demand a proof.[1]

At the risk of being charged with absurdity and betraying a pitiable absence of reflection, seeking proofs for the *PSR* is precisely what I intend to do in the first half of this chapter! This small act of rebellion against Schopenhauer does not imply that the issue he raises here is not worthy of addressing. On the contrary, there is a seeming absurdity and circularity to the question: is there a sufficient reason for the truth of the principle of sufficient reason?[2] What I would wish to claim, though, is that it is just that—i.e. it is a *seeming* absurdity and circularity, and not an actual one.[3] Schopenhauer is not correct to state that "whoever requires a proof for this principle […] already assumes thereby that it is true," as it is of course common to demand proofs for all sorts of claims without such demands being construed as tacit acceptance of the *PSR*. In this respect, requiring a proof of the *PSR* is no different to requiring a proof for any other putative truth.

With this in mind, let us turn to our arguments in favour of the *PSR*. There are three to be examined: first, an argument courtesy of Leibniz, who offers us definitions followed by a demonstration of the supposed fact that, whatever is has a sufficient reason; second, a Wolffian argument, originally presented in his *Ontologia*, to the

---

[1] Schopenhauer (1974), pp. 32-3. Schopenhauer's choice of language in the above passage is somewhat ambiguous: is he simply opposed to the idea that we should *demand* a proof for the *PSR*; or is he making the altogether stronger claim that even to *seek* a proof is misguided? We can certainly do the latter without being committed to the former, although Schopenhauer fails to acknowledge this. Still, from the context in which Schopenhauer writes—a brief diatribe against the "fruitless attempts" of others to prove the *PSR*—and his explicit statement earlier in the same chapter that he "[hopes] to show […] that the principle in general cannot be proved," we can be left in no doubt that it is the stronger of the two claims that Schopenhauer wishes to advance.

[2] A similar point is made by Leibniz (1956a: p. 717) who, after describing the *PSR* as a principle that states the "want" of a sufficient reason for any truth or event, asks: "Is this a principle that wants to be proved?"

[3] Aside from the seeming absurdity and circularity of providing a proof for the *PSR*, a further explanation for the comparative lack of arguments for the principle is that its truth is considered so evident that it simply requires no defence. Pruss (2011: pp. 13-4) writes: "[T]hose philosophers who accept the PSR typically do so because they take it to be *self-evident* and hence only in need of refinement and defence from attempts at disproof, but not in need of proof."

effect that the falsity of the *PSR* would entail a contradiction, and hence the *PSR* must be true; and finally, a contemporary line of defence presented by Pruss, who argues that the *PSR* cannot be denied except on pain of irrationality. Each argument will be followed by critical discussion.

2.1.1 Leibniz's Argument

Leibniz provides the very first argument for the *PSR* which, accordingly, we shall examine first. This argument makes its initial appearance in the early 1670s, relatively early in Leibniz's philosophical career, and is restated a couple of times hence, the final time in 1716.[4] We are first provided with a couple of definitions, and following on from these, a demonstration of the truth of the *PSR* is presented:

Definition 1: A sufficient reason is that which is such that if it is posited the thing is

Definition 2: A requirement is that which is such that if it is not posited the thing is not

Demonstration:

(1) Whatever is, has all its requirements (for, by *def. 2*, if one of them is not posited, the thing is not).
(2) If all its requirements are posited, the thing is (for if it is not, it will be kept from being by the lack of something, that is, a requirement).
Therefore all the requirements are a sufficient reason by *def. 1*.
Therefore whatever is has a sufficient reason.[5]

Leibniz's strategy, then, is to define the terms 'sufficient reason' and 'requirement,' and employ these definitions to support both his premises and conclusions. Another way of phrasing Leibniz's conclusion would be to say: if and only if a thing has all its requirements—or, what amounts to the same thing, has a *sufficient reason*—then it exists. Leibniz supports this conclusion by the use of seemingly acceptable

[4] The initial statement of the argument is to be found in *Sämtliche Schriften und Briefe* (1986: VI, ii, 483); it is then repeated in the same source (VI, iii, 118); and finally it makes an abridged appearance in Leibniz's last letter to the redoubtable Samuel Clarke (V, 18).
[5] I understand Leibniz's use of "is" in the statement "whatever is has a sufficient reason" to have the broadest application. In other words, I take it not only to mean that the *existence* of any particular thing requires explanation, but also that changes in state of any thing requires explanation.

definitions and premises that together lead ineluctably to the *PSR*. Nevertheless, many have found problems with both Leibniz's definitions and his premises, and we shall consider some of these now.

A first criticism to mention—and one which will be examined in greater detail when considering arguments against the *PSR*—comes from Hume. This criticism questions the truth of premise (1): Hume's contention is that, while Leibniz correctly defines a 'requirement' as anything that, if it is not posited, renders the 'thing' in question non-existent, he groundlessly assumes that all 'things' must in fact *have* requirements. So, while any existent thing must possess all of its requirements for being *if it has any* (and the mere definition of 'requirement' is a testament to this truth), this does not entail—and indeed it is not the case, says Hume—that all things actually have requirements.

How might the claim that all things have requirements be undermined, according to Hume? The answer is: by questioning the veracity of the law of causality. Hume argues that, contrary to what is generally assumed, we can easily conceive of a thing coming into being without any cause at all: " 'Tis a general maxim in philosophy, that *whatever begins to exist, must have a cause of existence*. This is commonly taken for granted in all reasonings, without any proof given or demanded […] But we shall discover in this maxim no mark of any such intuitive certainty."[6] And if a thing can come into existence in the absence of any cause, that thing plausibly lacks any requirements for its existence, and will therefore provide a counterexample to premise (1).

In fact, it is not that Hume provides a concrete example of any particular thing that causelessly exists; but that providing one is not necessary for his argument, as far as Hume sees it at least. Instead, what Hume wishes to claim is that it is possible to *conceive* of a thing's coming to exist causelessly; and anything that can be conceived of without contradiction or absurdity is "therefore incapable of being refuted by any reasoning from mere ideas; without which 'tis impossible to demonstrate the necessity of a cause."[7]

---

[6] Hume (1972), pp. 78-9.
[7] *Ibid.*, p. 80.

As mentioned, Hume's critique of the *PSR* will be exposited and critiqued at greater length when we come to examine arguments against the *PSR*. For now, it is sufficient to note that one possible line of attack on Leibniz's argument available to the *PSR* sceptic is to deny premise (1) on the grounds that it falsely assumes that all things have requirements.

A second criticism of the argument, rather unsurprisingly, questions premise (2). The claim advanced in premise (2) that, if all the requirements of a thing are posited then that thing exists, is called into question by evidence from contemporary physics. Again, this line of attack will be examined in greater detail in the succeeding section ('Arguments against the *PSR*'), although the criticism can be sketched here. According to many—perhaps the majority—of quantum physicists, experimental results from this field establish the fact that there is a certain degree of randomness to some quantum events. And if this is correct, if experimental results do indeed establish the fact that there is randomness at the quantum level, then it must surely be correct to say that there are some instances in which all the requirements for some thing are present, yet that thing fails to occur.

To offer a concrete example in order to make the objection to Leibniz's second premise clearer, we need look no further than the famous Schroedinger's Cat thought experiment.[8] According to this scenario, devised by Schroedinger in 1935, a cat is to be placed in a steel chamber together with a Geiger counter and a tiny amount of radioactive substance. In the course of one hour, there is a 50% chance of one of the atoms from the radioactive substance decaying. If an atom does decay, then the counter tube will discharge, which will release a hammer that will in turn shatter a flask of hydrocyanic acid, thus killing the cat.

What makes this scenario a potential counterexample to premise (2) is that, according to the *PSR* critic, *no explanation can be proffered as to why an atom*

[8] However, it should be noted that Schroedinger himself would not necessarily consider his thought experiment to be a counter-example to Leibniz's second premise. This is because the purpose of Schroedinger's thought experiment, as he saw it, was to provide a *reductio* of the prevailing Copenhagen interpretation of quantum mechanics, according to which the cat remains in what is known as a 'superposition' of states (i.e. both dead and alive) until an experimenter opens the steel chamber containing the cat. So, Schroedinger's thought experiment is designed to highlight the absurdity of describing the cat as remaining in a superposition of dead and alive states until the box is opened, although the example also serves as a vivid illustration of supposed indeterminacy at the quantum level.

either decays or fails to do so. All that can be said—and, indeed, all that there is to it—is that over the course of one hour there is a 50% chance of one of the atoms decaying, and a 50% chance that no atom will decay. This chance event at the quantum level translates into a chance event at the macroscopic level, with potentially catastrophic consequences for the unsuspecting cat. And the thought now is: it seems reasonable to characterise this situation as one in which all the requisites are in place to kill the cat, yet the cat might get lucky and live. Note, crucially, that the fortuitous feline's survival would be entirely providential, and not due to the absence of any requirement to kill it. In this way, Schroedinger's Cat presents a challenge to Leibniz's claim that, if all requirements for some thing are posited, then that thing is.

A further question is whether this objection should be taken as a challenge not just to premise (2), but also to the very definition of the term 'sufficient reason.' Pruss certainly thinks that this is a possibility, and thinks that by doing so one can thereby rescue the *PSR* itself. Pruss says of quantum systems such as that described in the Schroedinger's Cat thought experiment: "[we can] say that the system's initial state and its randomly causing A explain why A occurred."[9] On this interpretation of the thought experiment, then, all the requirements for the cat being killed provide the sufficient reason for the cat's being killed, although, contrary to Leibniz's definition of sufficient reason, these being posited do not entail that the cat *will* in fact be killed. Only after the fact can we know whether the cat will survive, since this will be determined by the chance event of an atom of radioactive substance decaying within the allotted timeframe. This defence of the *PSR* would be of no comfort to Leibniz, we can safely imagine, since it comes at too high a price: that is, it requires a fairly radical redefining of the concept of sufficient reason, one that admits of arbitrariness.

As mentioned earlier, this quantum mechanical objection will be examined in further detail when we look at arguments against the *PSR*. For now, and by way of summary, I want to add simply that Leibniz's argument is particularly useful since it spells out so clearly what the originator (and to those that believe, the discoverer!) of the *PSR* understood the principle to mean, and upon what assumptions it is founded.

[9] Pruss (2011), p. 168.

Both Leibniz's assumptions and definitions have been questioned, and we shall revisit these soon in the section on 'Arguments against the *PSR*' to see whether either his understanding of the *PSR* was faulty, or indeed whether the *PSR* is just plain false. Before we do, though, there are a couple more arguments to consider in favour of the *PSR*, the following one propounded by Wolff, the most eminent German philosopher to follow Leibniz.

2.1.2 Wolff's Argument

In §70 of his *Ontologia*, Wolff provides the following statement of what he is setting out to prove:

"Nothing exists without a sufficient reason for why it exists rather than does not exist. That is, if something is posited to exist, something must also be posited that explains why the first thing exists rather than does not exist."[10]

We can see from this quote that Wolff takes an epistemic reading of the *PSR*, according to which not only do all things have reasons for existing, but these reasons must be knowable by us. Note how this stands in contrast to Leibniz who, more circumspectly, qualified his statement that all things have reasons by admitting: "we cannot know these reasons in most cases."[11]

Nevertheless, Wolff's argument does not depend on understanding the *PSR* epistemically—it can just as profitably be used in defence of Leibniz's non-epistemic interpretation. Here is the argument as stated by Wolff:

(1) Either (i) nothing exists without a sufficient reason for why it exists rather than does not exist, or else (ii) something can exist without a sufficient reason for why it exists rather than does not exist (§53).

(2) If (ii), then some *A* exists without a sufficient reason for why it exists rather than does not exist (§56).

Therefore nothing is to be posited that explains why *A* exists: *A is admitted to exist because nothing is assumed to exist*.

[10] Wolff (1736), pp. 47-9.
[11] Leibniz (1956a), p. 646.

> *But* this is absurd (§69). Therefore, nothing exists without a sufficient reason; and if something is posited to exist, something else must be assumed that explains why that thing exists.[12]

The crucial step in this argument, it is widely considered, is the conclusion preceding the *reductio*, which is drawn from the two premises. In fact, Wolff draws two conclusions: while the sentences appear, on the face of it, to be articulating the same point in slightly different words, an equivocation is in fact revealed on analysis. The first—and incontrovertibly correct—conclusion is that, if some *A* is posited which lacks a sufficient reason for why it exists (as the *PSR* critic is bound to maintain), then "nothing is to be posited that explains why *A* exists." The second conclusion is the part that all of Wolff's critics have seized on as false, which is: "*A is admitted to exist because nothing is assumed to exist.*" This amounts to an equivocation since Wolff is using the word 'nothing' to mean different things in these two sentences, the first time legitimately and the second time not so. The first time, Wolff correctly concludes that, if the *PSR* is false then there is at least some thing (*A*) for which nothing can be posited that would explain its existence. The second time Wolff uses the term 'nothing', he claims that this nothing is (absurdly) what the *PSR* critic must posit in order to explain *A*'s existence; but of course the *PSR* critic can justly counter that this is a flagrant misunderstanding of what he means by saying that "nothing can be posited" to explain a certain thing's existence. He means, of course, that *no thing* can be posited, and not that nothing must be considered to be something.

In brief then, Wolff purports to reveal a contradiction that follows from the assumption that something exists without a sufficient reason: 'nothing' must absurdly be assumed to exist. Since this flouts the *Principle of Contradiction* (the *PoC*), the conclusion (which is the *PSR*) is claimed to follow.

In summary of Wolff's argument, there are two main points that I wish to note. Firstly, it is perplexing that Wolff should present such a bad argument. The seemingly wilful misuse of the notion of 'nothing' admits of no ready explanation, and Hume's assessment of the argument as "sophistical" is surely correct.[13] Still, perhaps we can understand why Wolff might construct such an argument if we

[12] Wolff (1736), p. 47.
[13] A judgment Kant appears to borrow when he concludes: "It is not difficult to escape the sophistry of the argument." (*NE* 1:398)

consider the context in which it was created. For Wolff, the issue of philosophical first principles was a pressing one, and he was keen in particular to resolve the issue of the relation between the *PoC* and the *PSR*, the two great principles handed down to him from Leibniz. And so the second point to note is that, in constructing his argument for the *PSR*, Wolff felt that he had found a means not only of demonstrating his conclusion, but also of showing how the *PSR* follows from the *PoC*, thereby neatly solving the problem of the hierarchy of principles. Now, with the failure of Wolff's argument, we are left with the question of what to conclude regarding both the relation between and the veracity of these two principles. Perhaps the failure of Wolff's argument indicates that the *PSR* cannot be derived from the *PoC*, and that both principles should simply be considered as axiomatic and not amenable to proofs.[14] Perhaps the most modest and therefore reasonable conclusion would simply be that the jury is out on whether the *PSR* is derivable from the *PoC*—Wolff's attempt was unsuccessful; but this does not preclude the possibility of future success on this score. Or perhaps, as Hume would argue, it is evidence (one piece among many) that the *PSR* is false. Certainly this latter conclusion would be far too hasty a one to draw yet, especially given that we have one more argument to examine in favour of the *PSR*. This final argument is suggested by Pruss, a modern-day *PSR* advocate.

2.1.3 Argument from Rationality

Pruss presents an argument for the *PSR* from rationality. It is extremely straightforward and is often only presented, says Pruss, as a straw man argument by the *PSR* critic for the purposes of being swiftly refuted. Nevertheless, Pruss thinks that there is more to the argument than is commonly suspected, and he goes on to defend it after presenting it. I shall follow suit here by stating the argument and then offering a defence. The argument runs as follows:

(1) The *PSR* says reality is rational, and denying the *PSR* is to admit to some degree of irrationality in reality.
(2) It is irrational to suppose reality to be in any degree irrational.

[14] This line of thinking would certainly appeal to Schopenhauer, who as we saw argued that the *PSR* is an innate principle that cannot be further analysed without descending into absurdity.

Therefore it is irrational to deny the *PSR*.[15]

The argument as presented is valid. The first premise is not really contestable either: it is hard to think of a clearer statement of a belief in the all-encompassing rationality of our world than the *PSR*, since it affirms as an exceptionless fact that nothing can be true in the absence of a sufficient reason. The burden must then fall on the second premise, which does indeed require considerable justification.

The first and most obvious thing to say about the second premise is that the *PSR* sceptic might well claim that there is nothing *prima facie* irrational about supposing that reality is not fully rational. The belief in the (partial) irrationality of reality might be compared to the belief that some individual (Kim Jong Il, say) is irrational: just as it is not irrational to claim that Kim Jong Il is irrational (on the contrary, he manifestly *is*), so is it not irrational to claim that reality is irrational.

While the *PSR* critic must surely accept the fact that it need not be *prima facie* irrational to think that reality is not always rational, I think the analogy is sufficiently *dis*analogous to allow the *PSR* proponent to further motivate the second premise. The purpose of the analogy was to show that it is not in all instances irrational to judge something to be irrational: just as it is *prima facie* rational to judge Kim Jong Il to be irrational, so too is it rational to judge reality itself not to be—at least not at all times—rational. Accepting this point, however, it is still open to the *PSR* proponent to provide reasons for thinking that, on reflection, it is irrational to suppose reality to be irrational.

A first consideration is that the irrationality imputed to persons such as Kim Jong Il and the irrationality imputed to reality as a whole are in fact two very different things. To say that a person is irrational can mean any number of things: that they are sometimes unpredictable; that they act in ways that go against the interests of themselves or others; or that their actions are often capricious or ill considered. It only takes a moment's reflection to note that reality cannot sensibly be considered to be irrational in *all* of the above ways: reality, not being a person, has no will and thus cannot act at all, either in ways that help or hinder its own or others' interests; similarly, lacking personhood, irrationality cannot be imputed to reality in the sense

[15] Pruss (2011), p. 249. I have adapted Pruss's own rendering of the argument slightly in order to ensure its validity.

of it acting capriciously or in an ill-considered manner (although for poetic purposes we occasionally talk as though it can, as when we bemoan being 'dealt a cruel hand by fate' and so forth).

Nevertheless, while different, there is overlap between the two usages, and this overlap consists in the notion of *unpredictability*. The *PSR* sceptic will want to claim that, just as certain people's behaviour can be irrational in the sense of being unpredictable, so too is reality as a whole irrational in just this sense. But in fact the claim is stronger than that: it is not just that reality is considered by the *PSR* sceptic to be unpredictable in *practice*, but rather that it is unpredictable in *principle*. That is, the unpredictability of reality is not merely a reflection of some epistemic weakness on our part—it is not a consequence of our inability to discern the connections between things, as it were—but is instead a reflection of the ontological fact that reality does not always conform to laws. The *PSR* proponent can hardly deny that aspects of reality are, at least at the present moment, beyond our capacity to fully explain or predict, with quantum mechanical events providing one such example of our epistemic limitations. But the *PSR* proponent will deny what the sceptic urges, that from this and other examples it should further be inferred that reality is irrational in the sense that there are some facts that are not amenable to explanation even to an omniscient mind. So here we come to the crux of the debate between the *PSR* proponent and sceptic: while the former affirms that all things are in principle explicable even if not in practice, the latter believes that there are at least some things which are explicable neither in principle nor in practice.

What can we say in favour of the *PSR* proponent's assertion that all things are at least in principle explicable, and that it is irrational to suppose otherwise? A first point to make is to suggest that an attitude of assuming that all things are explicable given enough inquiry is one that is bound to prove more beneficial to science and to making new discoveries. If we believe, on the other hand, that it is the case that some facts of nature simply have no explanation, such an attitude will surely be prejudicial to the search for truth and explanation. In fact, if we assume that the *PSR* is false, then the worry might be that this could open the door to a profoundly anti-rational scepticism.

While it is true that, from the point of view of scientific discovery it is better to err on the side of assuming that a given fact or event is explicable, this is not argument enough to support the conclusion that it is positively irrational to suppose that our universe might not always conform to law-like behaviour. The additional claim that assuming the *PSR* to be false leaves the way open for an anti-rational scepticism is unfair also, since the *PSR* sceptic can say that there are principled reasons for thinking that in specific instances—such as regarding events at the quantum level - the universe defies rationality.[16] In summary, if the *PSR* proponent's argument for the second premise rests solely on the claim that assuming the truth of the *PSR* will prove beneficial to science, then the *PSR* sceptic is entitled to respond that, while this may be true, it in no way implies that all facts and events do in fact conform to reason. The *PSR* proponent's point here is a merely pragmatic one, counselling an attitude of open-mindedness regarding the possibility of uncovering the universe's secrets. Acceptance of this approach in no way compels the *PSR* sceptic to forsake the belief that our universe will at times defy rationality.

More can be said, however, in defence of premise (2), besides urging the pragmatic benefits of assuming that our universe conforms to reason. A further important point to note is that we typically have a *desire* to seek explanations. The fact that we possess this desire indicates (it could be argued) that there are explanations to be sought, since we do not normally form desires for things that do not exist.[17] In response to this, of course, the *PSR* sceptic would certainly counter that a desire for explanation can simply be seen as an evolutionarily useful trait, and not as evidence that there are in all instances explanations to be uncovered. Again, the desire for explanation is of pragmatic benefit, since in instances in which there are explanations to be found, having the desire to uncover these is a prerequisite to actually uncovering them. In this way, our desire is for explanation can itself be explained by the *PSR* sceptic without them needing to concede that all things without exception must yield an explanation.

As has been shown, the second premise of this argument is certainly its most contentious feature, and this discussion unfortunately leaves the debate concerning

[16] The principled reason might be something along the lines of: after extensive observations and experiments, certain facts or events seem to resist all rational explanation.
[17] For further discussion of this point, see §72, §73, and §74 of Wolff (1736), and Gurr (1959), pp. 41-2.

its truth unresolved. However, we can see more clearly now a central bone of contention between the *PSR* proponent and sceptic, which is the issue of whether the unpredictability of certain features of our world should be characterised as epistemic or ontic. That is, do apparent violations of the *PSR* simply reflect our ignorance regarding the workings of our world as the *PSR* proponent insists (and thus are these violations *merely* apparent), or do they rather provide evidence that sufficient reasons are not to be found in all instances as the *PSR* critic believes? This is a question which we will be in a better position to answer at the end of the following section, which deals with arguments against the *PSR*. Certain supposed non-deterministic events, such as those posited by "standard interpretation" of quantum phenomena, will be considered as counter-examples to the claim that the *PSR* is an exceptionless truth. The first argument to consider in detail against the *PSR*, however, is Hume's, which we will revisit having already briefly assayed his challenge to Leibniz's claim that whatever is has all its requirements.

2.2 Arguments Against the PSR

Three more arguments are to be considered in this section, this time for the opposite conclusion: that the *PSR* is false. The first argument, both historically and in the order in which we shall examine them, is Hume's, which relies on the idea that there is nothing contradictory in the notion of a thing existing or something happening without a reason or cause. A second argument, a good deal more contemporary this time, comes from van Inwagen, who argues that the *PSR* is false since assuming its truth involves accepting an absurd consequence: namely, that everything happens of necessity and thus that all modal distinctions are an illusion. The final argument can actually be considered as a cluster of arguments possessing the same form. The idea is that there is some actual event or phenomenon in the world that can be shown to lack a sufficient reason for its happening. Contenders for this status of 'Reasonless Event or Phenomenon' include quantum phenomena, certain freely willed human actions, and even the universe in its entirety. In this section, it is quantum phenomena that will be under the microscope.

2.2.1 Hume's Argument

Hume's sceptical view of causality has proved hugely influential. In fact, regarding Hume's attack on common sense metaphysical views of causation, no lesser person than Kant was to declare: "I freely admit that it was the remembrance of David Hume which, many years ago, first interrupted my dogmatic slumber and gave my investigations in the field of speculative philosophy a completely different direction."[18] Still, not all have been so quick to praise Hume's work in this area, with Schopenhauer dismissively claiming: "Everyone at once feels the fallacy [of Hume's argument]."[19] The task before us now is to set out Hume's argument, whose conclusion is that the common belief in the necessity of a cause to all things is false, and then to judge whether to side with Kant or Schopenhauer in our assessment of it.

Hume's challenge to the *PSR* lies in the claim that we can easily conceive of a thing coming into being without any cause at all. He asserts:

"'Tis a general maxim in philosophy, that *whatever begins to exist, must have a cause of existence*. This is commonly taken for granted in all reasonings, without any proof given or demanded […] But we shall discover in this maxim no mark of any such intuitive certainty."[20]

Here we have Hume's conclusion: that, contrary to popular wisdom, it is not the case that whatever begins to exist requires a cause to explain its existence. And how does he justify this conclusion? By arguing that the proposition that 'whatever has a beginning has also a cause of existence' is incapable of being demonstrated, as follows:

We can never demonstrate the necessity of a cause to every new existence, or new modification of existence, without showing at the same time the impossibility there is, that anything can ever exist without some productive principle; and where the latter proposition cannot be proved, we must despair of ever being able to prove the former.[21]

Hume goes on to argue that we cannot prove the latter:

[A]s the ideas of cause and effect are evidently distinct, 'twill be easy for us to conceive any object to be non-existent this moment, and existent the next, without conjoining to it the distinct idea of a cause or productive principle. The separation, therefore, of the idea of a cause from that of a beginning of

[18] Kant (2004) 4, 260; 10.
[19] Schopenhauer (1974), p. 29.
[20] Hume (1972), pp. 81-2.
[21] *Ibid.*, p. 82.

existence, is plainly possible for the imagination; and consequently the actual separation of these objects is so far possible, that it implies no contradiction or absurdity; and is therefore incapable of being refuted by any reasoning from mere ideas; without which 'tis impossible to demonstrate the necessity of a cause[22]

We can state Hume's argument as follows:

(1) If causes are demonstrably necessary, then we must be able to demonstrate the impossibility of anything coming to exist without some productive principle.
(2) It is easy to conceive of some object being non-existent one moment and existent the next.
(3) Whatever we can conceive of—those things that imply no contradiction or absurdity—is possible.

Therefore it is possible for something to come to exist without some productive principle (*contra* premise (1) consequent).

Therefore it is not the case that whatever begins to exist must have a cause of existence (*contra* premise (1) antecedent).[23]

Premises (1) and (3) of Hume's argument will be questioned, beginning with Premise (1). Hume asserts that, if causes are demonstrably necessary, then it must be possible to demonstrate the *impossibility* of anything existing without a cause. In fact, there is nothing objectionable about this assertion: it is true that, were it possible to demonstrate the necessity of a cause, then this would entail the impossibility of anything existing in the absence of a cause. The objection, rather, lies in Hume's demand that the *PSR* proponent must be able to *demonstrate* the necessity of a cause: in response we can observe that, even though we remain unable to provide any demonstration as to why it is the case, causes might very well be universally necessary. If this is correct, then it must equally be the case that, despite being unable to demonstrate the impossibility of anything coming to exist in the absence of a cause, such an occurrence is as a matter of fact impossible. The accusation here, then, is that Hume sets the bar too high for the *PSR* proponent: he

[22] *Ibid.*, p. 82.
[23] For the sake of a humorous aside I shall now quote the ancient Greek philosopher Chrysippus (in Cicero (1991: pp. 20-1)), whose dire prognostications on the consequences of uncaused movement show him to be the antithesis of Hume: "Nothing exists or has come into being in the cosmos without a cause. The universe will be disrupted and disintegrate into pieces and cease to be a unity functioning as a single system, if any uncaused movement is introduced into it."

demands that the necessity of a cause be demonstrable, and thus (conversely) that the impossibility of anything existing without a cause must be demonstrable. The *PSR* proponent can reasonably claim that Hume's first conclusion (that it is possible for something to come to exist without some productive principle) does not follow from his denial of the consequent of the first premise, by consideration of the fact that perhaps not all things that are necessary are at the same time demonstrable. It is true that all things require causes in order to exist, the *PSR* proponent will claim, although this cannot be demonstrated.

In fact, a way to turn the first premise on its head might be to highlight the difficulty of demonstrating that any particular thing *lacks* a cause, rather than having to demonstrate that all things must *have* a cause, a difficulty to which Hume seems to be oblivious. Since there is no reason to think that causes cannot on occasion be spatially or temporally remote events, it seems that the entire universe would need to be scoured before it could be declared beyond any doubt that some particular event or thing lacked a cause. If, therefore, there must always remain a possibility that any particular event has a cause, then Hume cannot conclude that it is not the case that whatever begins to exist must have a cause of existence.[24]

The next challenge to Hume's argument concerns the claim made in premise (3), which is that whatever we can conceive of is possible. Kant rejects this claim, writing in his *Lectures on Metaphysics*:

> The Principle of Contradiction is the criterion of possibility, and one can also say of truth, only not a sufficient one but rather a necessary condition. If that is so, can I reverse: what does not contradict itself, is not impossible? That is false, for otherwise all fantasies that do not contradict themselves would be possibilities.[25]

Kant is saying that, while abiding by the *PoC* is a necessary condition for possibilities, it is not also a sufficient condition for them (although it is both necessary and sufficient for conceptions). So, while whatever does not contradict itself is conceivable, it is *not* the case that such a thing is in all cases possible. Recall

[24] In fact, however, nowhere does Hume declare that the *PSR* is undoubtedly false—all we know for certain was he believed it is not provably true. Nevertheless, given his insistence on the possibility of causeless things and events, it seems natural to conclude that he believes that, on occasion, such causeless things and events do actually occur. And if he does believe this, then he is guilty of a double standard in requiring a proof for the necessity of a cause from the *PSR* proponent, while at the same time being unable himself to provide a proof that any particular thing lacks a cause.

[25] Kant (1997), 29:791.

also Spinoza's assertion that there are some things that imply no contradiction or absurdity, yet their existence is not possible on account of the fact that it would not follow from the workings of corporeal nature. This, too, amounts to a denial of the equivalence of conceivability and possibility. Armed with this distinction, the *PSR* proponent can accept that causeless existence is *conceivable* (since such an event would not violate the *PoC*), but nevertheless deny the inference made in premise (3): that causeless existence is *possible*. Another way to put the same point is to say that not all things that are necessary are analytically so: some things are necessary because of the concepts involved (the oft-cited example that all bachelors are unmarried springs to mind), whereas other things are necessary despite their negation not involving a contradiction. The fact that everything necessarily has a cause falls into this latter category, and if we can only elaborate on this by saying that causes follow from the order of corporeal nature, or that causes are metaphysically necessary, while this is not entirely satisfactory it is nonetheless perfectly clear what is being stated. It is also more plausible, I would argue, than the claim that the bounds of the possible are identical to the bounds of the conceivable.

To conclude our remarks on Hume's argument, the first point I wish to urge is that Hume is wrong to assume that the idea of the necessity of a cause arises purely through observation and experience. It is this that motivates him to assert that the *PSR* proponent must demonstrate that causes are necessary—a feat that (it must be confessed) cannot be achieved. However, the *PSR* proponent need not be able to demonstrate the necessity of a cause, since it is more correct to consider this idea of the necessity of a cause to be an *a priori* assumption that is borne out by experience. Further, Hume fails to convincingly illustrate the possibility of causeless existence: on the contrary, I have argued, causeless existence is certainly impossible, even if it is (arguably) conceivable.

In short, the *PSR* proponent need not be able to demonstrate the *impossibility* of something coming to exist causelessly (since they do not claim that they can demonstrate the necessity of a cause); and Hume fails to demonstrate the *possibility* of something coming to exist causelessly. Premises (1) and (3) thus let the argument down. Hume's challenge is valuable, though, since it helps us to clarify why we might hold the *PSR* to be true. Hume claims: "Since it is not from knowledge or any scientific reasoning, that we derive the opinion of the necessity of a cause to every

new production, that opinion must necessarily arise from observation and experience."[26] However, the reality is that people hold the *PSR* to be true not solely on account of observation and experience (although observation and experience do indeed corroborate the principle), but rather because it is a judgement that we seem to be hardwired to hold, unless, as seems to have been the case with Hume, excessive philosophical speculation unseats it.

2.2.2 Van Inwagen's Modal Argument

The second argument against the *PSR* to consider brings us up to the twentieth century. It is presented by van Inwagen, who charges that the *PSR* has the "absurd consequence" of requiring the collapsing of all modal distinctions.[27] The argument runs thus: first, van Inwagen notes that the *PSR* requires that any contingent state of affairs must have a sufficient reason for its obtaining (that is, the sufficient reason for that state of affairs must lie in some other state of affairs that is necessary). Then, he asks us to imagine that P is the conjunction of all contingently true propositions into a single proposition. Now, according to the *PSR*, there must exist some state of affairs S that is a sufficient reason for P. We are then faced with a dilemma, since S must be either contingent or necessary; but it seems that it cannot be either. It cannot be contingent, because then it would be a part of P (since P is the conjunction of all contingently true propositions); and it cannot be necessary, since this would render P necessary also (since P necessarily follows from S). Van Inwagen concludes: "Since S cannot be either necessary or contingent, it cannot exist, and the *PSR* is false."[28]

---

[26] Hume (1972), p. 80.

[27] Van Inwagen is not the first person to present such an 'argument from contingency'. Rowe (1975: pp. 99-100) advances a similar argument against the *PSR*, which runs as follows:

(1) The *PSR* implies that every state of affairs has a reason either within itself or in some other state of affairs
(2) There are contingent states of affairs
(3) If there are contingent states of affairs then there is some state of affairs for which there is no reason
Therefore the *PSR* is false

I present van Inwagen's argument here, however, since his has received more attention in the philosophical literature and is perhaps the slightly more straightforward presentation of the two. Bennett (1984: p. 115) also offers a very brief presentation of the same argument, but adds nothing of interest over and above Rowe's and van Inwagen's presentations.

[28] Van Inwagen (1983), p. 203.

(1) If the *PSR* is true, then everything happens of necessity.
(2) It is not the case (on pain of absurdity) that everything happens of necessity.
Therefore the *PSR* is false.[29]

Both of van Inwagen's premises can be challenged. Beginning with the first premise recall how, in response to Leibniz's argument for the *PSR*, there was a worry that quantum events such as the one described in the Schroedinger's Cat thought experiment might present a challenge to the claim that, when all the requirements of a thing are posited then that thing is. In that particular instance, all the requirements might reasonably be said to have been in place for an atom of radioactive substance to decay within the course of an hour, yet an atom might nonetheless fail to decay (if the cat gets lucky). This then led us to consider Pruss's suggestion, which was that Leibniz's definition of 'sufficient reason' might be at fault, and that, once this definition had been amended, the *PSR* could be rescued. Rather than defining the term 'sufficient reason' in such a way that its presence entails the existence of the explanandum, Pruss favours loosening the connection between explanans and explanandum somewhat, in order to allow for the possibility of indeterminism, whether this be indeterminism at the quantum level, in human action, or in some other domain. Pruss suggests:

> We could understand the phrase *randomly caused* to be analogous to *freely chose* and say that the system's initial state and its randomly causing *A* explains why *A* occurred. Here, we can invoke the principle that when we have given the causes of an event, we have explained why the event occurred.[30]

If we were to reinterpret the term 'sufficient reason' as Pruss recommends, then this would allow for the possibility of indeterminism and thereby enable us to avoid the spectre of necessitarianism (that is, the thesis that everything happens of necessity). Van Inwagen's argument would fail, since the entailment relation posited in premise (1) would no longer hold. Another consideration in favour of this approach is that Leibniz himself seemed to vacillate between endorsing the definition for 'sufficient reason' given in the preceding argument, and adopting an altogether less strict

[29] Van Inwagen presents much the same argument in his book *Metaphysics* (2009: pp. 119-122). The only difference in the treatment of the argument this second time is that he begins with the assumption that there are contingent states of affairs, and from this he deduces the falsity of the *PSR*.
[30] Pruss (2011), p. 168.

understanding of the term, according to which a sufficient reason "merely inclines and does not truly necessitate."[31]

However, the cost of reinterpreting the notion of a sufficient reason along these lines is fairly high, I would argue, not least because of the difficulties of explaining in what sense a reason can be said to be 'sufficient' for some thing or event if it does not in fact entail that that thing or event obtains.[32] There is also a question mark over Leibniz's motivation for redefining the term 'sufficient reason' in some of his later writings so as to suggest that, on occasion, a sufficient reason "merely inclines": rather than being moved by philosophical conviction, it could well be an expedient measure to avoid being charged with the heretical opinion that God lacks free will.[33] In conclusion, while the *PSR* proponent could choose to adopt the above response to van Inwagen's argument, I think a better approach is to accept the entailment relation in premise (1) and focus instead on undermining premise (2).

According to premise (2), the collapsing of all modal distinctions entailed by adherence to the *PSR* is an "absurd consequence," one that renders acceptance of the *PSR* itself absurd. But I argue that the *PSR* proponent can counter that what van Inwagen's argument demonstrates is not the falsity of the *PSR*, but rather the impossibility of there being contingent truths. So, rather than seeing the collapsing of all modal distinctions as an absurd *consequence* of commitment to the *PSR*, the

---

[31] Leibniz (1956a), p. 226.

[32] On the question of how to define the word 'sufficient', we see Kant protesting against the ambiguity of this designation in his *New Elucidations*, and preferring instead to talk of grounds that 'determine'. In Kant's own words: "…the expression 'sufficient' is ambiguous, for it is not immediately clear how much is sufficient." (*NE* I:393) Using the term 'determining', by contrast, is supposed to make it clear that the ground in question entails that the thing or event obtains. As for the term 'ground', it is likely that Kant also favours this over 'reason' out of a desire to avoid ambiguity or downright error. Kant's problem with the term 'reason' is that he perceives its use as encouraging an epistemological reading of the *PSR*, according to which a reason is that which enables us to know something. The term 'ground,' meanwhile, fits more readily with an ontological reading of the *PSR*: in other words, the ground is that which makes it the case that something is. Evidence that it is indeed for these reasons (or on these grounds!) that Kant opted for the term 'ground' over 'reason' comes from his critique of Wolff, whom he admonishes for understanding reasons "in terms of that by reference to which it is possible *to understand why* something should be rather than not be." (*NE* I:392, my italics) To sum up, I agree with Kant both that: (a) reasons/grounds should be understood as *determining* the thing or event in question, and that; (b) reasons/grounds should be understood to mean the reason/ground why something is *simpliciter*, as opposed to the reason/ground for *understanding or knowing that* something is. However, I also think that the term 'sufficient reason' is not as ambiguous as Kant fears, and that it has enough historical standing to warrant its retention.

[33] The urge to carve out his own philosophical niche, distinct from Spinoza, must also have been an influencing factor on Leibniz. And the infamy attaching to Spinoza on account of his necessitarianism makes it quite understandable that Leibniz should want to avoid a similar fate—holding such views did not just make one unpopular, it was potentially dangerous given the intolerance of the time.

PSR proponent can counter that assuming the existence of contingent states of affairs is an unwarranted *supposition*. What the argument seems to come down to, then, is the question of which is the more reasonable supposition: that (as van Inwagen asserts) since there exist contingent truths the *PSR* must be false; or that (as the *PSR* proponent asserts) since the *PSR* is true there must exist only necessary truths.

I would argue that the *PSR* proponent's position is more defensible, for two reasons. First, we can understand van Inwagen's unwillingness to abandon the idea of the existence of contingent truths as a consequence of the importance of contingency as a feature of our psychology. When making decisions, for example, it is necessary to deliberate as if the future exhibits genuine openness. Also, given our extremely limited knowledge of things, it is hardly surprising that some things will appear to us as brute facts. In other words, to say that something *might* happen, or *could* happen, merely reflects a lack of knowledge on the part of the speaker.[34]

Second, necessitarianism makes for a much neater ontology in that it bypasses worries about the ontological status of possible worlds.[35] If contingency is allowed for, then we are faced with the problem of deciding how it is that possible worlds are supposed to exist, whereas necessitarianism faces no comparable problem: there is one world, the actual one, and there is simply no such thing as a possible world. A further worry is that, if there are possible worlds, then in virtue of what fact is this possible world the actual world? Perhaps the difficulty in answering such a question would not trouble van Inwagen: he might feel disinclined to supply a reason in this case, since as a critic of the *PSR* he openly doubts that reasons always exist. Nevertheless, to those who have an intuitive inclination towards belief in the *PSR*, the lack of potential explanation is a problem, and van Inwagen's bald assertion that necessitarianism is absurd will do little to persuade the *PSR* proponent to concede the argument.

---

[34] This is an idea endorsed by history's most famous necessitarian, Benedict Spinoza. In his *Ethics* (1992: Pr. 33, Sch.1, p. 54), Spinoza asserts: "If we do not know whether the essence of a thing involves a contradiction, or if, knowing full well that its essence does not involve a contradiction, we still cannot make any certain judgment as to its existence because the chain of causes is hidden from us, then that thing cannot ever appear to us either as necessary or as impossible. So we term it either 'contingent' or 'possible.'"

[35] We must judge, for example, whether some version of modal realism is the correct view, whereby all possible worlds are as real as the actual world. (See Lewis (1986) for an exposition of modal realism.)

Instead of hoping to persuade the *PSR* proponent through the use of more conceptual arguments such as those of Hume and van Inwagen, perhaps only hard evidence of events that demonstrably lack any sufficient reason would do the trick. We have briefly examined quantum events already, and we shall return to them now to see if they yield counterexamples to the claim that the *PSR* is true.

2.2.3 Non-Deterministic Events

Let us remind ourselves of how those two arch-rationalists Spinoza and Leibniz understand the *PSR*. Spinoza says: "From a given determinate cause, the effect follows necessarily. Conversely, if there is no determinate cause, it is impossible for an effect to follow."[36] And Leibniz offers the following definition of 'sufficient reason,' as we saw earlier: "A sufficient reason is that which is such that if it is posited the thing is."[37] It seems that both of these statements amount to an endorsement of determinism, and so of course rule out the possibility that there could be any non-deterministic events. We can state the situation as follows: the notion that (a) the *PSR* is uniformly true, allied to the understanding that (b) the term 'sufficient reason' is to be defined in such a way that sufficient reasons necessitate their explanandum, entails the truth of determinism. And therefore, if some non-deterministic event can be discovered, this will provide a counterexample to the claim that the *PSR* is true.

Quantum theory is one such candidate for providing a counterexample to the claim that the *PSR* is true. There are two major respects in which quantum theory is thought to represent a departure from classical physics: first, it does not seem to require locality of causation; second, it is apparently non-deterministic. It is this second feature of quantum theory that is of interest to us in this context—the apparent non-locality of causes presents no threat to the *PSR* since the *PSR* proponent need not stipulate that all causes be local.

We are now in a position to state the argument against the *PSR*:

(1) If the *PSR* is true, then determinism is the case.

[36] Spinoza (1992), Ia, 3.
[37] Leibniz (1986), VI, ii, 483.

(2) Quantum theory provides evidence that some events are non-deterministic.

Therefore, the *PSR* is false.

The argument structure recalls van Inwagen's modal argument: the first premise makes a claim about what the truth of the *PSR* would entail, and the next premise claims that we have reason to think that the thing entailed by the *PSR* is false. Thus, it is concluded, the *PSR* must itself be false. Still, the present argument is somewhat stronger, I would argue, for a couple of reasons. First, it is less contentious to argue that the *PSR* entails determinism than that it entails necessitarianism (although both of these claims are, I think, correct). Second, and more crucially, the present argument cites particular events—namely, quantum ones—that purportedly demonstrate the falsity of the consequent (i.e. determinism), whereas van Inwagen simply proclaims that to accept the consequent (i.e. necessitarianism) would be absurd.

In discussing responses to van Inwagen's modal argument, I dismissed the idea of denying that the PSR entails necessitarianism. Of course, if we accept that the *PSR* entails necessitarianism then *a fortiori* we must accept that it entails determinism, and so premise (1) is secure.[38] Instead, our efforts should be focused on undermining premise (2) with its claim that there is evidence for indeterminism from quantum theory. It must be admitted that some august names in quantum physics have believed the evidence tells against determinism, as illustrated by Heisenberg's quote: "Since all experiments are subjected to the laws of quantum mechanics, the invalidity of the law of causality is definitively proved by quantum mechanics."[39] Nevertheless, equally august names (not least Einstein) have argued that the evidence is far from conclusive. David Bohm is among those who have believed that the evidence for indeterminism at the quantum level is questionable, and he created his own "hidden variable" theory as a rival account to the "standard interpretation." According to "hidden variable" theories such as that of Bohm, "there is some additional factor (a hidden mechanism) such that once we discover and understand this factor, we would be able to predict the observed behaviour […] with

[38] See Pruss (2011: pp. 103-13), however, for an attempt to make the case that the *PSR* does not entail determinism.

[39] Heisenberg (1927), p. 197.

certainty."[40] The essence of all "hidden variable" theories, then, is the idea that the indeterminism we perceive at the quantum level is a consequence of epistemic limitations. As such, the indeterminism perceived is simply that—a mere *perception*—and so we are not entitled to infer that events at the quantum level really are non-deterministic.[41]

This approach to evaluating our understanding of quantum phenomena has a precedent in another area of science. Statistical mechanics is a branch of physics that applies probability theory to the study of the thermodynamic behaviour of systems comprised of a large number of particles. Classical mechanics, on the other hand, encompasses a broader spectrum, concerned as it is with discovering the set of physical laws that describe the motion of bodies under the action of a system of forces. Statistical mechanics is thus a probabilistic science that exists within the framework of classical mechanics. Bohm hoped and predicted that the statistical description in quantum theory would come to take within a completed quantum theory "an approximately analogous position to statistical mechanics within the framework of classical mechanics."[42] This was Einstein's hope also.

It might be suggested that the key question to ask regarding the relative merits of the "standard interpretation" and Bohm's "hidden variable" theory is: which is superior in terms of its ability to make correct predictions? Unfortunately, answering this question does not help us to adjudicate between the two since they are empirically evenly matched, with both able to offer reliable predictions. Despite their equivalence in terms of predictive power, however, in other respects the Bohmian theory is certainly the superior of the two. For example, some experiments and experimental issues are dealt with much better by the Bohmian theory, examples of which include experiments on quantum chaos, scattering theory, dwell and

[40] Bishop (2002), p. 117.
[41] Bohmian theory can be seen as a response to Heisenberg's "Uncertainty Principle" which states that, for certain pairs of physical properties, the more precisely we know what the one is, the less precisely are we able to measure the other. The Bohmian theorist's response is to declare that there is always a matter of fact about both the position and momentum of any particle—each and every particle has a well-defined trajectory. However, it is added, observers have limited knowledge as to what this trajectory is (and thus a limited knowledge of the particle's position and momentum). It is the lack of knowledge of the particle's trajectory that accounts for the uncertainty relation. To put the statement differently, while the particle's position and velocity can only be known statistically, this does *not* imply that the particle's trajectory is indeterminate.
[42] See Dürr *et al*. (1992), p. 269.

tunnelling times, and escape times and positions.[43] Interestingly, there do not seem to be any examples to illustrate the converse: in other words, there are no examples to illustrate how in some instances the "standard interpretation" handles experiments and experimental issues better than Bohmian theory. This consideration of Bohmian theory's superiority in certain experimental settings is thus a real point in its favour.

It is worth noting also that Schroedinger's famous thought experiment was *not* devised for the purpose of demonstrating that quantum indeterminism could result in macro-level indeterminism. Rather, it was intended as a *reductio* of the "standard interpretation" of quantum physics, according to which the cat, prior to the experimenter opening the steel chamber in which it is contained, should be described as existing in a "quantum superposition" of dead-and-alive states.[44] Since the "standard interpretation" seems to absurdly imply that the cat remains in some sort of limbo state between death and life until being observed by the experimenter, we must conclude that the understanding it affords of quantum mechanics is incomplete.[45]

Now that Bohmian theory has been mooted as a viable, deterministic alternative to the "standard interpretation," let us consider a couple of hurdles it must still overcome. First, we have the following worry presented by Pruss:

> If all we want is to see that it is *conceptually* possible to have a deterministic theory that gives the same observational results as quantum mechanics, it is easy to see that this can be done. For instance, take a neo-Leibnizian theory that says that every point of space is a monad, and this monad has encoded within it a list of all the events that will happen throughout time at that point, and through an internal causal process it goes deterministically through these events as time passes.[46]

---

[43] For experiments on quantum chaos, see Cushing (2000); for experiments on scattering theory, see Dürr *et al*. (2000); for experiments looking at dwell and tunnelling times, see Leavens (1996); and for experiments on escape times and positions, see Daumer *et al*. (1997).

[44] Elaborating on the intention behind his thought experiment, Schroedinger (1936: p. 328) writes: "It is typical of these cases [i.e. cases such as the one outlined in Schoedinger's Cat thought experiment] that an indeterminacy originally restricted to the atomic domain becomes transformed into macroscopic indeterminacy, which can then be *resolved* by direct observation. That prevents us from so naively accepting as valid a "blurred model" for representing reality." Schroedinger's intention is thus clearly to question the possibility of indeterminism rather than affirm it.

[45] And, it must be added, Bohmian theory offers a sorely needed response to this conceptual flaw in the "standard interpretation." As the highly respected physicist John Stewart Bell (1987: p. 160) notes: "Bohm showed explicitly how parameters could indeed be introduced […] with the help of which the indeterminate description [of the "standard interpretation"] could be transformed into a determinate one. More importantly, in my opinion, the subjectivity of the orthodox version, the necessary reference to the 'observer', could be eliminated."

[46] Pruss (2011), p. 167.

The challenge is that, if one is determined enough to deny that determinism's false (no pun intended) then this can always be done since, strictly speaking, no amount of evidence can falsify it.[47] Even a profoundly ludicrous deterministic theory, as Leibniz's monadic theory assuredly is, can be argued to be consistent with observations from quantum mechanics. Pruss wishes to argue, then, that the defender of Bohmian theory is in fact in an analogous position to the neo-Leibnizian—they are backing a theory purely out of a dogged determination to affirm determinism in the face of all countervailing evidence. Determinism's unfalsifiability, it is charged, is being guilefully exploited as a licence to ignore the reasoned arguments of others.

To my mind, this charge does not stick. It is true, as Pruss notes, that it is conceptually possible to devise any old deterministic theory that will be concordant with the observational results of quantum mechanics, and if someone held to a neo-Leibnizian theory such as is sketched above then that could be quite justifiably be dismissed as being unreasonable (although not provably false). However—and as has been discussed already—there are good reasons to prefer Bohmian theory over the "standard interpretation," as not only do the two have equal predictive power, but Bohmian theory is superior in many experimental setups, and does not face the same troubling subjectivity that the "standard interpretation" faces with its necessary reference to an observer.[48] The analogy is therefore unfair, since it likens a discredited theory with no predictive power to one which any dispassionate observer would agree is very much reputable.

A second difficulty to consider is as follows: if Bohmian theory is a genuine contender for the theory that best describes what happens on the quantum level, why is the "standard interpretation" still just that: i.e. the interpretation which scientists consider the standard, and to which they most regularly refer? As Bell observes: "the Bohr interpretation [i.e. the "standard interpretation"], in its more pragmatic, less metaphysical forms remains the "working philosophy" for the average physicist."[49]

---

[47] It is always possible, for example, that any putatively causeless thing or event may in fact have either a hidden or a (spatially or temporally) remote cause.

[48] Bell (1987: p. 160), in a critique of the tendency to accept the "standard interpretation" unquestioningly, makes the point that we need not uncritically accept this position with all of its troublesome consequences: "vagueness, subjectivity, and indeterminism, are not forced on us by experimental facts, but by deliberate theoretical choice."

[49] Bell (1987), p. 189.

In response to this, the first thing to say is to highlight that the "standard interpretation" is described as a "working philosophy" for most physicists. In other words, it would probably be misleading to regard its use as representing an endorsement of the philosophy that underlies it. It is used simply because it works. Another point to note is that, while the "standard interpretation" remains the working philosophy for most physicists, there has been increasing interest in what have been termed "many worlds" theories.[50] Such theories also come under the banner of "hidden variable" theories, and so also typically involve an attempt to reconcile quantum phenomena and determinism. There is not the space for further discussion of these here, and I think that all "many worlds" theories face serious conceptual problems that make championing them an unattractive proposition.[51] Nevertheless, their increasing popularity over the last decade or so provides evidence that serious interest in "hidden variable" theories remains. Yet another response to this difficulty is to suggest that it might not be wise to be too respectful of scientific orthodoxy. Evidence for this assertion comes from the fact that in 1932, a paper was published by a certain von Neumann, claiming to prove that all "hidden variable" theories were impossible. It was generally decreed that von Neumann's reasoning was unimpeachable, and that all "hidden variable" theories were indeed impossible. However, three years later, Grete Hermann discovered that von Neumann's result was flawed. To the scientific community's discredit, this fact went unnoticed for a further fifty years. The moral of the story seems to be that current scientific orthodoxy does not always provide a reliable guide to truth.

A final point to note on the issue on the likelihood of Bohmian theory becoming orthodoxy is this: quantum mechanics is comparatively in its infancy, and so the fact that the dominant theory at the moment is non-deterministic should not be afforded too much significance. Whatever one's opinion on the likelihood of quantum theory eventually being reconciled with determinism, there must surely be a consensus that our current understanding of quantum phenomena admits of improvement.[52] Quite simply, we should *expect* at this stage not to have a full understanding of events at

[50] Deutsch's (1997) provides an overview of the "many worlds" interpretation of quantum physics from its leading exponent.
[51] See Wallace *et al*. (2012) for an excellent collection of essays examining the pros and cons of "many worlds" theories.
[52] As Bell (1987: p. 201) urges: "To admit things not visible to the gross creatures that we are is, in my opinion, to show a decent humility, and not just a lamentable addiction to metaphysics."

the quantum level. And the further expectation of the Bohmian theorist is that an improved understanding can and will vindicate their belief that determinism and quantum phenomena go hand-in-hand.

The crux of the quantum theory debate can be summarised by Hodgson's quote: "The key question is whether to understand the nature of [the probability in quantum theory] as epistemic or ontic."[53] My answer is that it is epistemic. The supposed indeterminism that quantum phenomena exhibit is a persistent fiction which, given time, will be revealed as such.

2.3 Conclusion

Six arguments have been examined, three for and three against the *PSR*. Leibniz's argument for the *PSR* is excellent as a jumping-off point, as its premises illustrate very clearly two key potential avenues of objection, enabling us to map the arguments against the *PSR* that we have considered onto it. First, as Hume objects, it is wrong to assume that all things have requirements. Second, it is possible that something might fail to happen despite all its requirements being posited, as the "standard interpretation" of quantum theory would suggest. However, in response to the first objection, I have argued that Hume fails establish his conclusion, which is that requirements (or, what amounts to the same thing, causes) are not always necessary; and to the second objection, I suggested that Bohmian theory is superior to the "standard interpretation," and its adoption would allow us to maintain that, as Leibniz asserted, if all the requirements for some event are posited then that event will of necessity occur.

Both of these avenues of objection, against Leibniz's first and second premise premises, counsel a loosening of the bonds between the existence of a thing on the one hand, and the explanation for its existence on the other. Van Inwagen, it can be presumed, would object to one or other – or indeed both – of Leibniz's premises in order to allow for contingency and to avoid the "absurdity" of necessitarianism. Without van Inwagen offering any clearer idea of how contingency enters the picture, however, I argued that the *PSR* proponent is entitled to reject the supposition

[53] Hodgson (2002), p. 99.

of contingency and embrace necessitarianism instead. Granted, it is a consequence of the *PSR*; but it is not an absurd one.

A second argument for the *PSR*, presented by Wolff, turned out to be ineffective due to an embarrassingly transparent equivocation. Nonetheless, the argument was useful in that it encouraged consideration of whether the *PSR* might be derived from some yet more fundamental principle such as the *PoC*. The answer appears to be: if it is so derived, it is not obvious how it is so.

The final argument in favour of the *PSR* was Pruss's "Argument from Rationality." The key premise was the second one, which stated that belief in the irrationality of our world is itself irrational. The contention—that our desire for explanations might indicate that, in all cases, there are in fact explanations to be had—turned out to be too contentious; the conclusion was that, on the basis of this argument alone, no conclusion could be reached. However, discussion again proved fruitful as it helped to clarify the central issue at stake between *PSR* proponent and critic. This issue is: are events or phenomena that appear random, or that are beyond our present understanding, evidence for the falsity of the *PSR*? To this question, the proponent would respond that there is no such event or phenomenon, however ostensibly inexplicable, which provides a counterexample to the *PSR*. The critic, on the other hand, answers that, at least in some instances, events or phenomena do provide counterexamples to the *PSR*.

While the arguments remain inconclusive, I submit that the case for the truth of the *PSR* appears stronger at present than the case against. Leibniz's argument spells out most clearly the case for the *PSR*, and in the absence of proof that his premises are faulty—since, against the first premise, Hume fails to demonstrate that anything might come to exist without requirements, while, against the second premise, our knowledge of quantum physics does not provide reason for thinking that something might fail to happen despite all of its requirements being posited—I am inclined to accept the conclusion, and suggest we adopt a presumption in its favour.

In terms of what follows from the truth of Leibniz's and Spinoza's understandings of the *PSR*, these were seen to entail the truth of determinism: everything has a sufficient reason, and sufficient reasons necessitate their explananda. Consequently, any demonstrably non-deterministic event (as certain quantum events were held to

be, for instance) would provide a counterexample to the *PSR*, and it is the possibility of finding such an event that remains the *PSR* sceptic's best hope for challenging the principle. Thus, in our next chapter, we shall be looking at just such a potential counterexample, at a theory that deals with a domain much closer to everyday reality than the rarefied world of quantum mechanics: the domain is that of human action, and the theory is that of libertarian free will.

3

Is Libertarianism a Counterexample to the *PSR*?

Towards the end of the previous chapter, it was noted that the *PSR* critic's best hope of establishing the principle's falsity would be to identify some event, phenomenon, or theory that acts as a counterexample to the *PSR*. In this chapter, the thesis of libertarianism will be considered as just such a counterexample.

The thesis of libertarianism holds that free will is logically incompatible with a deterministic universe and that, since we have free will, it follows that determinism is false. In particular, libertarians maintain that *human actions* must be undetermined if we are to have the free will that they believe we do. Where classical quantum mechanics threatened the truth of the *PSR* by virtue of positing undetermined events at the quantum level, libertarianism does so by positing undetermined events in the realm of human action instead. It is these undetermined events in the realm of human action, libertarians assert, that demonstrate the falsity of the *PSR*, since, as has been argued, the principle entails the truth of determinism.[1]

In order to establish whether the libertarian challenge to the *PSR* has merit, we will examine two particular libertarian theories—Ginet's and Kane's. From our evaluation of these, we can extrapolate to make conclusions concerning the success of libertarian theories in general as counterexamples to the *PSR*.

[1] In fact, there is a strong case, I believe, for thinking that the *PSR* entails the truth of necessitarianism. This point will not be pressed here, however, since in the present dialectical situation it merely needs to be established that determinism follows from the *PSR*.

3.1 Ginet's Libertarianism

As with all libertarians, Ginet argues that free will is not compatible with determinism. What he understands by 'free will' is the following:

> By freedom of the will is meant *freedom of action*. I have freedom of action at any given moment if more than one alternative action is then *open to me*. […] Two or more alternatives are *open to me* at a given moment if which of them I do next is entirely up to my choice at that moment: Nothing that exists up to that moment stands in the way of my doing next any one of the alternatives.[2]

So, for Ginet, freedom of will is the same thing as freedom of action, and possessing freedom of action amounts to the ability to act in one of a variety of ways at any given moment. Another way of phrasing this would be to say that, according to Ginet, free will requires that the state of the world at any given moment (a state that includes not just all facts about the external world but also all facts about the agent's inner state) must not determine how that agent decides to act. Freedom of will—and so by extension freedom of action—amounts to an agent's having alternative possibilities open to them, and an agent freely wills or acts when this willing or action is not determined, whether by facts about themselves or the world.[3] This insistence that *all* free acts must be ones for which an agent has alternative possibilities means that Ginet's libertarianism is more demanding than some other libertarian theories: for instance Kane (whose theory we will be examining shortly) does not insist on this point; rather, he argues that free will merely requires that *not all* an agent's acts are determined.[4]

What does Ginet think that an action is? Put simply, all actions can be classified as being one of two things: either an action is a "simple mental act," or else it is a "voluntary exertion."[5] The latter of these, Ginet explains, is in all cases preceded by the former, because all voluntary exertions must begin with volition, and volition is always a causally simple mental action.

---

[2] Ginet (1990), p. 90.
[3] Ginet is very clear on the necessity of alternative possibilities for an action to be free. As a definition of a free action, Ginet (1990: p. 17) offers the following: "By a *free action* I mean one such that until the time of its occurrence the agent had it in her power to perform some alternative action (or to be inactive) instead."
[4] See Fischer *et al*. (2007), pp. 5-7.
[5] Ginet (1990), pp. 14-15.

Ginet goes on to describe simple mental acts as having "an *actish* phenomenal quality," by which he means to convey that it feels "as if I directly make [these simple mental acts] occur, as if I directly determine [them]."[6] Ginet is keen to stress the importance in this description of the qualifying phrase "as if," as he does not want to give the impression of there being a causal relation between the agent on the one hand and the mental occurrence on the other. If there were such a causal relation then Ginet's libertarianism could be classified as an 'agent-causal' view, according to which an agent is a persisting substance, an uncaused cause of their free decisions. But Ginet rejects agent-causal accounts of libertarianism, primarily because he believes that such accounts cannot explain what they need to explain: "It cannot, for instance explain [a mental act's] timing. The mere fact that I was there cannot explain why this mental act occurred just when it did rather than earlier or later, when I was also there."[7] Similarly, Ginet rejects the idea that simple mental acts are 'event-causal,' on the grounds that the simple mental act "fails to have a sufficiently complex structure."[8] In other words, if event-causation were the correct analysis of human action we would expect simple mental acts to appear much more complex than they do in fact appear.

Agent-causal and event-causal views thus both fail to do justice to the nature of simple mental acts, according to Ginet, who views such acts as "counterexamples to the thesis that acting is causing."[9] In place of these, Ginet presents his own 'non-causal' account of action, which he sums up as follows:

<<*S*'s *V*-ing at *t*>> designates an action if and only if either (i) it designates a simple mental occurrence that had the actish phenomenal quality or (ii) it designates *S*'s causing something, that is, an event consisting in something's being caused by an action of *S*'s[10]

So on Ginet's view, any action—whether it be (i) a simple mental act or (ii) a voluntary exertion—must be uncaused by virtue of the fact that all actions begin with an uncaused simple mental act. Nothing up until the moment of action determines how an agent will choose to act—it is instead simply a spontaneous

[6] *Ibid.*, p. 13.
[7] *Ibid*., pp. 13-4.
[8] *Ibid.*, p. 14.
[9] *Ibid*., p. 14.
[10] *Ibid*., p. 15.

choice on their part. Here is how Ginet elaborates on what it is for an action to be uncaused:

> It seems evident to me that, given that an action was uncaused, all its agent had to do to make it the case that she performed that action was to perform it. If my deciding to vote for the motion, for example, was uncaused (that is, nothing other than me determined or made it the case that I raised my hand), then I made it the case that I raised my hand simply by raising my hand.[11]

In summary, we have seen that Ginet is a libertarian on account of the fact that he believes that the existence of free will requires the falsity of determinism, and specifically that it requires that there be indeterminacy in human action. In contrast to some libertarians, however, Ginet further believes that for *any* action to be free it must be the case that, up until the time of its occurrence, the agent had it within their power to perform some alternative action or perhaps no action at all. Ginet further argues that all actions are in fact free in this sense, due to the fact that they are uncaused—as he says, all an agent must do to make it the case that they perform an action is to perform it.

So while Ginet believes in indeterminism in the sphere of human action as do all libertarians, he differs with his agent-causal and event-causal peers as to how this indeterminism occurs. While agent-causal libertarians propose that the agent is an abiding substance with the ability to initiate causal chains without being determined by the state of the world immediately prior to that instant, and event-causal libertarians locate indeterminacy within the agent at some stage during the decision-making process, Ginet's non-causal libertarianism is perhaps simpler: his assertion is that all actions, beginning as they do with an uncaused mental act, are themselves uncaused, and so of course remain undetermined up until the time the agent chooses to act.[12] It is "as if" the agent directly causes the simple mental act—although of course the agent is no substance on Ginet's view, and no allegedly freedom-compromising causation is involved in the process.

---

[11] Ginet (2007), p. 247.

[12] Describing Ginet's non-causal libertarianism as simple is not intended to be pejorative. In fact, Ginet himself dubs his non-causal account of human action "simple libertarianism" in his (1997) article "Freedom, Responsibility, and Agency."

With Ginet's non-causal libertarianism sketched, we can examine a couple of objections to his view. In evaluating these objections, our discussion of the *PSR* from the previous two chapters will prove relevant.

3.2 Criticisms of Ginet's Libertarianism

3.2.1 The Luck Objection

The first objection to consider can be termed the 'luck objection.' The basic thought is that, regarding the uncaused acts that Ginet posits, it must be concluded that it could only be a matter of pure luck as to how an agent chooses to act. The further conclusion to draw from this is that the agent cannot be responsible for any act that is subject to chance in the way that Ginet's uncaused actions appear to be. Ginet would respond that, far from being a *bar* to ascribing moral responsibility to agents, the uncaused actions he posits are instead a *prerequisite* for a world peopled with morally responsible agents. Note that this luck objection can be applied to any theory insofar as it posits indeterminism in human action, and thus to all libertarian theories: to all of these the objection can be levelled that indeterminism in human action cannot be posited without at the same time conceding that these actions are subject to luck, and, insofar as this is the case, the agent cannot therefore be held responsible for them.[13]

What does Ginet have to say in response to this objection? He responds that it is possible to draw a distinction between uncaused *events* on the one hand, and uncaused *actions* on the other, and that in drawing this distinction his critic's

[13] This objection has indeed been regularly levelled at the libertarian throughout the centuries of debate on the free will issue. For example, Hume (2007: sect. 8, pt. 1) asserts: "[L]iberty, when opposed to necessity, is the same thing as chance." A.J. Ayer (1954: p. 275) makes what is essentially the same point: "Either it is an accident that I choose to act as I do or it is not. If it is an accident, then it is merely a matter of chance that I did not choose otherwise; and if it is merely a matter of chance that I did not choose otherwise, it is surely irrational to hold me morally responsible for choosing as I did. But if it is not an accident that I choose to do one thing rather than another, then presumably there is some causal explanation of my choice: and in that case we are led back to determinism." Closer to the present day, we have Galen Strawson (1994: p. 7) echoing this line of thought in his claim that "it is absurd to suppose that indeterministic or random factors, for which one is *ex hypothesi* in no way responsible, can themselves contribute in any way to one's being truly morally responsible for how one is." Finally, Mele (1999: p. 99) argues that if nothing about an agent (e.g. their capacities, powers, states of mind, moral character etc.) causally determines how they will act, then their acting in one way rather than another is "just a matter of luck."

conclusion that undetermined actions must be subject to luck can be blocked. While an uncaused *event* (if there are any) could indeed be characterised as random, Ginet questions the legitimacy of characterising an uncaused *action* as random also, and will only concede them to be so if this characterisation does not entail that these actions are not "up to" the agent (that is to say, that they are not the agent's responsibility). In short, if by "random" we mean only that an event is not predictable from antecedent causes, then Ginet is happy to accept that the uncaused actions he postulates are "random" events; but he insists that "it is very far from evident […] that it cannot have been up to the agent whether a "random" action occurred and cannot have been the case that the agent made a "random" action occur."[14] As such, Ginet remains unconvinced that the 'luck objection' presents him with any difficulty in ascribing moral responsibility to agents for their uncaused actions.

The challenge Ginet lays down to the supporter of the 'luck objection' is thus to demonstrate how one of Ginet's postulated uncaused actions—one that is "random" in the sense that it is not even in principle predictable from antecedent causes—can be proved to be "random" in the further sense that its occurrence is (a) a matter of luck and (b) therefore not truly anyone's choice or responsibility. Taking up that challenge, van Inwagen offers a thought experiment designed to elicit from us the intuition that an uncaused action must be random in this sense. We are asked to suppose that a girl called Alice is faced with a choice between lying and telling the truth, and that when faced with this choice she decides to tell the truth. We are further asked to assume that this decision was undetermined. Following this unremarkable event, we must then suppose that God decides to cause the universe to revert to the precise state it was in one minute before Alice told the truth. When asked what will happen this second time, it seems that we can only say that she might lie or she might tell the truth. Van Inwagen continues:

> Suppose that God *a thousand times* caused the universe to revert to exactly the state it was in [one minute before Alice told the truth]. Suppose that [after a thousand replays] Alice has told the truth four hundred and ninety-three times and has lied five hundred and eight times. Is it not true that as we

[14] Ginet (2007), p. 248.

watch the number of replays increase, we shall become convinced that what will happen in the *next* replay is a matter of chance?[15]

Van Inwagen clearly presents this rhetorical question expecting the answer to be "yes": we *will* become convinced that what happens in the next replay is a matter of chance. Ginet, however, is not willing to accept this conclusion from van Inwagen's thought experiment. Ginet responds as follows:

If I contemplate just one Alice making an uncaused choice, I fail to see how the proposition that the choice was causally undetermined entails that it was random and not up to Alice which choice she made; and I quite fail to see how supposing there to be a great many duplicates of Alice in duplicate situations, sometimes making the same choice as Alice and sometimes making a different one, should make me any more inclined to think with regard to any one of these choices that its being undetermined entails that it is random and not up to its subject.[16]

Broadly speaking, I side with van Inwagen on the question of the implications of the above thought experiment. Since Alice's choice over whether or not to lie is causally undetermined, it seems that whether or not she lies on any given occasion can surely be considered a mere matter of chance; and consequently, it is at least debatable whether it would be fair to hold her morally responsible for her choice over whether or not to lie (although her supposed lack of moral responsibility in such a situation is much harder to establish, I think, than that her choice is subject to chance: this issue will be discussed shortly).

Dealing first with the issue of whether Alice—an agent with the capacity for uncaused action, living in a libertarian world of the type Ginet postulates—makes choices that are subject to luck, it seems to me obvious that she does. If we suppose that there are "duplicates" of Alice, as Ginet suggests, living in different yet identical worlds, and that some of these duplicates tell the truth as our original Alice did while others lie, how can we avoid the conclusion that whether or not any one of them lies is a matter of luck? Since Ginet is unwilling to concede that Alice's decision in the above scenario is subject to luck, perhaps we can make van Inwagen's conclusion more compelling by altering the scenario slightly: let us imagine instead that 99% of these identical Alices choose to tell the truth, while a mere 1% choose instead to lie; let us imagine further that telling this lie, whatever it is, has the catastrophic

[15] Van Inwagen (1983), p. 128.
[16] Ginet (2007), p. 249.

consequence of leading to the deaths of millions (telling the truth, meanwhile, had no deleterious effects, either for the 99% of Alices or for the populations of the worlds which they inhabit). Since, according to the hypothesis, all of these Alices were exactly alike in every respect, it seems not only *natural* to say of the 1% who lied and thus suffered the dreadful consequences of this that they were extremely unlucky, but in fact this admission seems completely *unavoidable*.

What this thought experiment also highlights, besides the fact that undetermined actions are by their nature subject to luck, is the role that the *PSR* plays in helping us reach this conclusion. If we take one of the 99% of worlds in which Alice tells the truth, and then take one of the 1% of worlds in which Alice lies, and then ask ourselves the question regarding these two worlds: "what is the reason that in this world Alice told the truth, whereas in this other world Alice lied?," it is clear that there can be no answer. There can be no sufficient reason why in one world Alice tells the truth while in another identical world she lies. And because there is no sufficient reason—because nothing fully explains why some Alices choose to lie while others choose to tell the truth—it is true to say that the 1% of Alices that choose to lie and must then suffer the consequences are terribly unlucky.

We saw how Ginet sought to respond to the 'luck objection' by drawing a distinction between uncaused events and uncaused actions, and then arguing that only the former are (a) subject to luck, and (b) not anyone's choice or responsibility. I think it is clear from the thought experiment developed from van Inwagen that this distinction simply does not hold, at least when it comes to the issue of luck. An uncaused action is one that is not determined by antecedent causes: as such, it lacks a sufficient reason for its occurrence; lacking a sufficient reason, whether it happens is subject to chance or luck, a fact that the above thought experiment inescapably leads us to conclude.

It is less obvious, however, that we can further conclude from the fact that uncaused actions must of necessity be subject to luck, that they cannot therefore be anyone's responsibility. In fact, it seems clear that we can construct a scenario in which we certainly would consider an agent responsible for their uncaused action. We can imagine an agent called Jim who, while on his morning route to work, comes across a man lying injured in the street. Since Jim is by nature a good man, he is 100%

likely to stop and see what is wrong (or to put the same point slightly differently, multiple Jims living in different yet identical worlds would all choose to stop). Further, upon finding that the injured man has lost a significant amount of blood, Jim is fully prepared to risk the ire of his boss by turning up late for work in order to call an ambulance to take this man to hospital. On top of that, there is an 85% chance that he will decide to give blood to this man, which would be not merely a commendable action but a supererogatory one. By contrast, an altogether less civic-minded gentleman called Jon, when faced with this same situation, might choose to stop only 10% of the time, while the remaining 90% of the time he would instead opt to continue on his way to work out of a selfish desire for financial advancement.[17]

The upshot of this thought experiment is that, if we concentrate on just one of the 85% of those Jims who take the heroic step of deciding to undergo a blood transfusion for this wounded stranger, it seems natural to suppose that he deserves credit for this act despite its being uncaused and subject to luck. Equally, it is natural to feel that any one of the 90% of Jons who fail to so much as stop for the injured man is deserving of censure, although once again this failure to act is uncaused and subject to luck. I think the moral to draw from this example is that, *as long as there is some relationship between a person's character and their actions, it is natural to think that they are (at least to some extent) morally responsible for these actions*. And since it is possible to sketch an account of uncaused actions without entirely severing the relationship between character and action, it would be a mistake to conclude from the mere fact that uncaused actions are subject to luck that uncaused actions cannot therefore be to some extent the agent's responsibility.[18]

What are we entitled to conclude from all this? One thing that cannot be concluded—or at least that I have not attempted to press for in the preceding discussion—is that the libertarian's metaphysical picture must be false. Perhaps if we were able to establish the premise that we definitely possess the free will required

[17] Jon is the manager of a hedge fund staffed exclusively by ex-Lehman Brothers employees.

[18] Note that the claim here is that it is *natural* to ascribe moral responsibility to someone whose acts are representative of their character; but this is not to say that it is also *correct*. Galen Strawson's argument, to be examined in the following chapter, argues that an agent should not even be considered morally responsible for actions that are expressive of their character, since they cannot be considered morally responsible for their character. For now, however, all I wish to note is that, while we can conclude that the occurrence or non-occurrence of undetermined actions is subject to luck, it is much less obvious that the presence of indeterminism precludes the possibility of moral responsibility also.

for moral responsibility, and were then able to demonstrate that the 'luck objection' precludes the possibility of us having this free will, then this objection would be decisive. However, whether we do have the free will required for moral responsibility is an open question at this stage. At any rate, it seems natural to think (as our Jim and Jon thought experiment shows) that, even if the uncaused events Ginet posits are of little help in securing moral responsibility, at least they do not preclude it.

Still, while we cannot conclude on the strength of the above that the libertarian's metaphysical picture is plain wrong, that is not to say that the 'luck objection' is not damaging. On the contrary, it throws into question whether the libertarian's primary objective can be achieved: namely, to present a picture of human action that affords us a freedom and/or responsibility that we would otherwise lack in a world of determined action. The crucial question for Ginet and his fellow libertarians is: if—as the luck objection suggests—it is a mere matter of luck whether one undetermined action happens rather than another, how can positing undetermined actions (whether these are the uncaused actions that Ginet posits or some other type of indeterminism) add one iota either to our freedom, or to the responsibility for our actions that we are said to enjoy? The answer, I contend, is that Ginet does not and cannot explain this supposed connection.

In conclusion, the 'luck objection' reveals that if, like Ginet, you wish to argue for indeterminism, you must also of necessity be committed to denying the *PSR*; and you cannot deny the *PSR* without accepting that any given event to which the *PSR* does not apply is undetermined and hence subject to luck (and that is true as much of human actions as it is of events). The charge against Ginet and his fellow libertarian, then, is this: *there is nothing to be gained in terms of increased freedom or responsibility for one's acts by denying determinism*.

The libertarian's commitment to *PSR* denial has passed without comment throughout the discussion of this 'luck objection.' That is to say, the legitimacy of the libertarian's commitment to denying the *PSR* has not itself been queried, but instead an unacknowledged—and what must also be for the libertarian an unacceptable—consequence of its denial has been brought to light (which is that any action lacking a sufficient reason must be subject to luck). The next objection to consider, however,

does question the legitimacy of denying the *PSR*. If the objection is successful it will not merely demonstrate (as the 'luck objection' has) that the libertarian is unable to make the case for the necessity of denying determinism in order to secure the freedom and/or responsibility they believe us to have—rather, it will simply demonstrate the falsity of the libertarian metaphysical picture. Let us examine this second objection now, using Ginet's non-causal libertarianism once more as the target of our critique.

3.2.2 The Lack of Explanation Objection

The 'luck objection' will be revisited when we come to examine Kane's theory of libertarian free will. Now, however, let us direct a second criticism towards Ginet's libertarianism, which is that his account allows for no adequate explanation of human action, and that therefore his account must be wrong. While this objection is similar to the previous one insofar as both reveal unwelcome implications to Ginet's indeterminism, its focus is slightly different: the first objection does not undermine the libertarian's metaphysical picture in any way, but rather seeks to demonstrate that this metaphysical picture does not provide what the libertarian hopes and expects of it (i.e. the extra choice and control over one's actions that would otherwise be lacking in a deterministic world, and that makes possible true freedom of the will); the second objection, however, *does* seek to directly undermine the libertarian's metaphysical position. It does so by appealing to the truth of the *PSR*, whose truth is thus central to the argument.

Ginet himself gives consideration to a version of the 'lack of explanation objection,' first formulating the argument, and then attacking it. Here is a (slightly reworked) presentation of the argument he considers:

(1) Incompatibilism entails that (at least) some free actions are not determined by an antecedent state of the world.
(2) If an action is not determined by an antecedent state of the world, then it has (at best) a partial explanation in terms of its antecedents.
(3) All free actions have complete explanations in terms of their antecedents.
Therefore incompatibilism is wrong.[19]

[19] Ginet's (1990: p. 129) own formulation is as follows:

The success or failure of the argument hangs on premise (3), which is essentially a statement of the *PSR*. The question is: does Ginet have convincing evidence for its falsity? Ginet himself certainly believes he does, as he offers the following counterexample to the premise:

> When I cross my legs while listening to a lecture, that action (usually) has no explanation in terms of reasons for doing it that I had antecedently. I just spontaneously do it. A spontaneous action, not arising from any antecedent motive, can even be undertaken with a further intention that begins to exist just when the action does. For example, a bird catches a person's eye and, without having antecedently formed the intention to keep watching it, she moves her head when the bird moves in order to keep her eyes on it.[20]

It can hardly be doubted that both of these examples are fairly characterised by Ginet as examples of spontaneous action, just in the sense that neither the act of crossing one's legs nor of having one's attention caught by a bird in flight are in any way planned or premeditated. However, if we are supposed to infer from these examples that the acts in question lack explanation or causal antecedents, then I see no reason to accept this inference—and in fact I see quite good reasons to reject it. First, in the case of the bird catching a person's eye, it is patently not true that no explanation can be given in terms of causal antecedents. While the act of following the flight of a bird is spontaneous, unbidden, and capricious (in the sense that it cannot be predicted), it is very clear that it is an act with a causal story, which involves the flight of the bird and the subsequent passing of the bird into the person's field of vision. It is not an act that springs from nowhere and admits of no explanation. As for the first example, this is perhaps a more promising candidate for an act that truly lacks explanation in terms of its antecedents, since there is nothing external to the

---

(1) Incompatibilism entails that an action cannot be both free and determined by an antecedent state of the world.
(2) If an action is not determined by an antecedent state of the world, then it has no explanation in terms of its antecedents.
(3) But some free actions do have explanations in terms of their antecedents.
Therefore incompatibilism is wrong.

Premise (1) has been altered to reflect the fact that some incompatibilists (such as Kane) do not hold that all free actions must be undetermined. Ginet's premise (2) is also too categorical, in that the critic of libertarian need not claim undetermined acts have *no* explanation in terms of their antecendents, merely that any explanation they do have must be at best *partial*. Conversely, premise (3) is needs to be more categorical to capture the *PSR* proponent's intuition: it is not merely that *some* free actions have explanations in terms of their antecedents, but rather that *all* do. Notwithstanding these tweaks to Ginet's premises, the basic issue at stake remains the same, namely the status of the *PSR*.

[20] *Ibid.*, p. 130.

agent that can be said to cause—or even to be part of a causal explanation of—the act of leg-crossing. Additionally, we are all familiar with the experience of performing some spontaneous action or other, whether crossing our legs, tilting our head, or folding our arms, and it must be said that the phenomenology of such acts is that they just happen, unconsciously and without forethought.

So it is an easy step from this to concluding that acts of these kinds lack explanations in terms of their antecedents; but it is nevertheless a completely unwarranted one. The assumption is that, just because we are unaware of the processes by which something comes to be, it must be the case that there is no explanation for why it should come to be. The examples with which Ginet intends to demonstrate that some free actions lack explanations fail to do this, and so we are entitled to demand more in the way of evidence before accepting this claim.

Ginet believes he does have more evidence. In the context of an attack on the necessitarianism of J.S. Mill, who makes the case that psychological laws are every bit as unvarying as physical laws, Ginet argues that it is sometimes the case that "when the same set of conflicting motives recurs, a different one prevails from the one that prevailed earlier."[21] By way of example, he cites the conflict that often arises within him between his desire to watch football and an opposing desire to work, as well as the occasional conflict between his desire to get an early start on the day and a desire to sleep in a bit. Ginet concludes from these examples: "If there are laws of nature that explain why the prevailing motive does prevail in such cases, that explain this in terms of antecedents of the action, it is far from obvious what the contents of those laws are." Rather than viewing an agent's actions as subject to laws of nature, Ginet surmises: "The agent had the conflicting motives, and it seemed a tossup which one to satisfy."[22]

Regarding Ginet's criticism of Mill's necessitarianism, there is a disanalogy between psychological and physical laws that makes the dispute impossible to settle conclusively. To pick one of Mill's own examples, whereas it is obvious how we go about testing the claim that a heavier weight will always lift up a lighter weight when each is placed on opposite ends of a pair of scales, there is no comparable method of

[21] *Ibid.*, p. 134.
[22] *Ibid.*, p. 134.

demonstrating beyond dispute that, given identical antecedents the same choice will always be made.[23] It is this difficulty that enables Ginet to present apparent counterexamples to Mill's assertion that a prevailing motive would always prevail in any situation in which the antecedents were the same.

However, for the same reason that Mill has difficulty in proving his assertion that all human actions are antecedently determined, so too does Ginet struggle to convince us of his conclusion to the contrary. It is simply not possible to engineer two situations in which all the antecedents are truly identical, in order to provide any accurate test of the proposition that, given identical antecedents, identical actions will follow. The best that Ginet can offer us, then, is anecdotal evidence from his own experience: accounts of similar occasions when he has felt conflicted between two courses of action, and has subsequently chosen opposing courses on these similar occasions. As evidence goes, this is lamentably poor. It is poor because, while two situations may be very similar in many respects (so, to borrow Ginet's example, in both situations it is a Saturday afternoon, there is football on the television, and the agent in question has competing desires either to get on with some work or to watch the football), no two situations will ever be identical. And no matter how similar any two situations are, the determinist can always assert without fear of disproof that it is the small but significant differences between the two situations that explain the different outcome in each case. So while Ginet views his example as evidence in favour of libertarianism, the determinist is able to offer his own perfectly coherent explanation for what is going on—and one, crucially, that does not require rejecting the *PSR*.

In conclusion, while Mill inevitably fails to prove beyond dispute that all human actions are antecedently determined (inevitably, since it is impossible to devise an experimental setup to test the uniformity of extremely complex and comparatively

[23] Others have noted the impossibility of verifying Mill's (1872: pp. 449-50) claim that "volitions do, in point of fact, follow determinate moral antecedents with the same uniformity and […] the same certainty, as physical effects follow their physical causes." For example, Steward (2012: pp. 170-1) asks: "Why, for instance, am I currently still sitting here writing rather than making the cup of tea I have been dimly thinking about for the past hour or calling the bank, which is something I also need to do today? Some will no doubt be inclined to say that I must have wanted to sit here writing more than I wanted the cup of tea and more than I wanted to phone the bank, and that this is the explanation. But if this 'must have' can be justified, it can surely only be by appeal to a conception of desire-strength according to which it simply follows from the fact that I am still sitting here that this was my 'strongest' desire."

poorly-understood psychological laws with the same accuracy that we can with physical laws), Ginet's anecdotal evidence in favour of indeterminism of human action is wholly unconvincing. Given the presumption in favour of the *PSR*, and Ginet's failure to offer any convincing reason to abandon this presumption, the lack of explanation objection is, I think, successful against Ginet's libertarianism. While Ginet's responses to both of these criticisms have been unconvincing, that is not to say that some other libertarian theory might not be able to provide a satisfactory response to them. For this reason, let us turn now to Kane, who believes he can answer these same objections satisfactorily.

3.3 Kane's Libertarianism

The next libertarian theory of free will to consider is Kane's, who champions what can be called an "event-causal" account of libertarian free will. This is to say that Kane argues free actions are caused by events involving agents. What distinguishes Kane's view from those of the many compatibilists who would agree with this basic analysis of free actions is that he adds the condition that these events must, at least on occasion, be non-deterministically caused.

Central to Kane's account is the claim that there are two key requirements for free will, both of which Kane believes are in fact met, and neither of which is compatible with determinism. The first of these is the familiar requirement of "alternative possibilities" (AP). Kane writes: "Free will seems to require that *open alternatives* or *alternative possibilities* lie before us—a garden of forking paths—and it is "up to us" which of these alternatives we choose."[24] Our future can be represented as a "garden of forking paths" in the sense that it really is open to us to choose from the many possibilities that lie before us—and our choices are not always determined.[25] Such an image vividly illustrates the difference between libertarian theories on the one hand, all of which affirm the existence of AP, and deterministic theories on the other,

[24] Fischer *et al.* (2007), p. 14.
[25] This choice of image originates from a short story by Jorge Luis Borges entitled "The Garden of Forking Paths."

according to which one's life would be more accurately represented by an image of a path with no branches or forks.[26]

This first requirement of AP is, as I have noted, a feature of all libertarian theories, and is summed up in Ginet's insistence that, when acting, it is often the case that two or more alternatives are *open to one*, in the sense that "which of them I do next is entirely up to my choice at that moment."[27] The second requirement for free will, by contrast, is not a feature of all libertarian theories. Nevertheless, Kane insists that its importance for the free will issue is as great, if not greater than, the AP requirement. This second requirement is captured by Kane when he declares it imperative that "the *sources* or *origins* of our actions lie "in us" rather than in something else (such as the decrees of fate, the foreordaining acts of God, or antecedent causes and laws of nature)."[28] This requirement is dubbed by Kane the condition of "ultimate responsibility" (UR), and he offers this further encapsulation of what it means to be ultimately responsible for an act: "*To be ultimately responsible for an action, an agent must be responsible for anything that is a sufficient cause or motive for the action's occurring.*"[29]

As to the relationship between AP and UR, the two supposed requirements for free will, Kane explains that the latter condition of UR does not require us to have been able to do otherwise—that is, to have had AP—for each and every act carried out of our own free will. However, it does require us to have had AP on at least some occasions in the past. Further, Kane contends that the choices we have made when faced with these AP are ones that have had a substantial impact in forming our present character, and he dubs these choices "self-forming actions" (SFAs). Summarising the importance of SFAs for explaining how it is that we can be morally responsible for our acts, Kane says: "SFAs are only a subset of those acts in life for which we are ultimately responsible and which are done "of our own free will." But if *none* of the acts in our lifetimes were self-forming in this way, we would not be *ultimately* responsible for anything we did."[30]

---

[26] For more on 'forking path' arguments, see van Inwagen (2009); Fischer (1994); and Ekstrom (2000).
[27] Ginet (1990), p. 90.
[28] Fischer *et al*. (2007), p. 14.
[29] *Ibid*., p. 14.
[30] *Ibid.*, p. 15.

In summary, Kane, in common with Ginet and indeed all libertarians, argues for the necessity of AP. But to this Kane adds the further requirement of UR, according to which an agent can only be ultimately responsible for an act if they are responsible for anything that is a sufficient cause or motive for that act. Since the requirement of UR could not be met without positing the existence of undetermined SFAs (because otherwise our choices would follow from events in the past for which we are patently not ultimately responsible), it follows that UR entails AP. In positing and describing the relation between AP, UR, and SFAs, Kane believes he has advanced a plausible and attractive theory of human action which demonstrates not only that: (a) agents are able (on occasion) to do otherwise, but also that; (b) they can do so voluntarily, intentionally, and rationally.

3.4 Criticisms of Kane's Libertarianism

3.4.1 The Luck Objection

The same objections that were levelled at Ginet's libertarian theory will be reconsidered in the light of Kane's own libertarian solution to the free will problem. The aim of course is to find out whether the difficulties that Ginet's theory faces are merely due to features of his particular brand of libertarianism, or whether in fact there is some intrinsic flaw from which libertarian theories as a whole suffer.

Let us begin once more by applying the luck objection to Kane's theory. Before examining his response we must first spell out how the luck objection applies to Kane's libertarianism. We need to know: how does Kane's account of human action invite the charge that the actions he describes are unavoidably and unacceptably subject to luck? In answer, the luck objection applies to Kane's undetermined SFAs: it is these actions, Kane's critic will say, that are unavoidably and unacceptably subject to luck. The argument is that, since SFAs are supposed to lack sufficient cause or motive (which, being undetermined, they must), the resulting action cannot be characterised as being anything other than (at least in part) subject to luck. This is the same oft-rehearsed argument that we saw was articulated by Hume, A.J. Ayer, Galen Strawson, Mele, and others when the same objection was raised to Ginet's libertarianism.

Kane, like Ginet, is familiar with this criticism, and his response to it is to argue that there is more going on with SFAs than the luck objection credits. Kane's contention is that the occurrence of SFAs is not the product of *mere* chance (although he concedes that there is an element of chance to all SFAs), but rather that it is primarily an effort of the will which ensures that one outcome prevails over another. Since we are *prima facie* responsible for our own efforts of will, we must equally be responsible for the undetermined SFAs that occur largely as a result of our efforts of will. To illustrate the importance of efforts of will in the occurrence of SFAs, Kane presents an example which he considers typical of cases in which SFAs are performed. This action, Kane is keen to stress, is willed, voluntary, and allows for a freedom that compatibilists find themselves deprived of due to their denial of AP. He describes the scenario thus:

> [A businesswoman] is on her way to a very important meeting when she observes an assault taking place in an alley. An inner struggle ensues between her conscience, to stop and call for help, and her career ambitions, which tell her she cannot miss this meeting. She has to make an effort of will to overcome the temptation to go on. If she overcomes this temptation, it will be the result of her effort, but if she fails, it will be because she did not *allow* her effort to succeed […] When we, like the woman, decide in such circumstances, and the indeterminate efforts we are making become determinate choices, we *make* one set of competing reasons or motives prevail over the others then and there *by deciding*.[31]

This response may appear convincing at first glance—Kane is surely right, after all, to think that efforts of will are of crucial importance when we are faced with conflicting motives. What is more, we are all familiar with facing difficult dilemmas: situations in which we find ourselves presented with two opposing courses of action and must choose between them as best we can. To come to a decision in these circumstances we must perhaps wrestle with our consciences, or maybe consider very carefully the options that are laid before us (or both of those things). However, notwithstanding the undoubted importance either of efforts of will when faced with a conflict of motives, or of the fact that these scenarios of 'feeling torn' are familiar to us all, a crucial question for Kane remains, namely: what specific role does indeterminism play in this process of decision-making for SFAs? Answering this is not such an easy task, a fact that is made plain by Pereboom's comment on Kane's position:

---

[31] *Ibid.*, pp. 26-7.

> Consider Anne, the businesswoman, who can either decide to stop and help the assault victim, or can refrain from so deciding. The relevant causal conditions antecedent to this decision—agent-involving events, or, alternatively, states of the agent—would leave it open whether this decision will occur, and she has no further causal role in determining whether it does. I contend that with the causal role of the antecedent conditions already given, whether the decision occurs is not then settled by anything about the agent—whether it be states or events in which the agent is involved, or the agent herself.[32]

It seems that Kane cannot avoid the charge that undetermined SFAs are subject to luck, given that the relevant causal conditions prior to an agent's decision when performing SFAs leave the outcome of this decision open. In the case of the businesswoman, if the relevant causal conditions prior to her decision leave it open whether she will help the assault victim, and if she plays no further causal role in determining how this decision is made, then whether she helps appears to be a matter of luck. So to the question of what role indeterminism plays in the process of decision making for SFAs, the inescapable answer is that all it introduces is an element of luck.

In conclusion, Kane's libertarian account is every bit as vulnerable to the luck objection as Ginet's. Whether by deliberate tactical manoeuvre or mere artless oversight, Kane succeeds in obscuring the issue by stressing that, when it comes to SFAs, it is the agent's will that is of central importance. From considering one's own experience, everyone can agree that situations do sometimes arise in which one must make an important decision between two opposing courses of action. Further, it seems clear that, especially when a decision must be made between an easier, less moral course of action and a harder, more moral one, a strong effort of the will may be involved. But nothing about this account of action conflicts with the idea that these inner struggles in fact have a *determinate outcome*. That is, nothing about the phenomenology of decision-making in such cases gives us reason to believe that this process is not wholly determined.

Further, Kane does not provide—and cannot by the nature of the case provide—an explanation of how the undetermined SFAs could transform what would otherwise be a determined and therefore supposedly unfree process into a free one. Once again, the evidence is that the libertarian theories so far examined fail to establish a link between indeterminism and increased freedom. Importantly, this is not due to a quirk

[32] *Ibid.*, p. 193.

of the particular theories presented, but is rather due to the fact that the one feature that all libertarian theories share—an embracing of indeterminism in the sphere of human action—is a feature that adds nothing to a determinist metaphysic beyond an element of chance.[33]

3.4.2 The Lack of Explanation Objection

Perhaps the best that the libertarian could hope for in the wake of the luck objection would be to pin their faith on future libertarian theories being able to explain exactly how indeterminism can be posited in such a way that it avoids the objection. This is certainly van Inwagen's position, who confesses himself unable to understand "how free will works," but feels justified in affirming libertarianism owing to his strong belief that humans possess the free will required for moral responsibility, allied to his further conviction that such free will would be impossible in a determined world.[34]

But what if the libertarian's metaphysical picture suffers not just the problem of being unable to convincingly explain 'how free will works,' as the luck objection reveals? What if it is not just the link between indeterminism and free will that is in question but something more fundamental, namely, the viability of the libertarian's metaphysical picture itself? If the *PSR* is to be embraced, then this is indeed the situation in which libertarians find themselves since the *PSR* has been shown to preclude indeterminism in *any* sphere of life, including of course the sphere of human action. This is the thrust of the 'lack of explanation' objection: since nothing happens or exists without having a complete explanation (as per the *PSR*), any theory of human action must reflect this fact; but libertarian theories, by virtue of espousing indeterminism, deny this. Libertarian theories are therefore untenable.

---

[33] To state the objection more generally, we might say: since all libertarian theories must posit AP, and since there must necessarily lack a sufficient cause or reason as to which AP actually occur, we cannot escape the conclusion that these so-called "freely willed" libertarian actions, to the extent that they are undetermined, are subject to luck.

[34] Van Inwagen (1983: pp. 216-7) expresses his position as follows: "I have never pretended to understand "how free will works." If I knew I would tell you, but I don't know […] I have no liking for unresolved mysteries in philosophy. But it is no good trying to pretend that mysteries do not exist if they quite plainly do exist. Moreover, I prefer small mysteries to large mysteries." To van Inwagen's mind, then, it is a smaller mystery to imagine that an agent is able to choose non-deterministically between outcomes than it is to suppose either that moral responsibility simply does not exist, or that moral responsibility exists "even though no one ever has any choice about anything."

With the 'lack of explanation' charge restated, let us see whether Kane is able to provide a libertarian rebuttal where Ginet could not. A good place to start the libertarian defence is with the following question: to what problem is indeterminism in the realm of human action deemed to be the solution? One answer to this is that indeterminism allows for alternative possibilities (AP). Without AP there is no real choice, it is claimed, as our futures must already be mapped out in front of us. Rather than our life paths being akin to a "garden of forking paths," we must instead be condemned to treading an unalterable course through our lives.

As arresting as this visual metaphor admittedly is, and as disquieting as is the notion that, if determinism is true, then our fates have been decided long before we were around to have any say in the matter, we have already seen that libertarians are incapable of securing AP without incurring the unacceptable cost of admitting that an individual's choice of AP must be ascribed to luck.[35] If there are forks in the path of life then it seems that there must be a corresponding absence of complete explanation as to why any particular path was in fact chosen, and hence such choices are subject to luck (as opposed to being the outcome of a freely-willed, libertarian choice).

The argument that indeterminism is necessary since it allows for AP is therefore a non-starter if the luck objection succeeds, as I have argued it does. But Kane offers a further reason for embracing indeterminism, and this rests on his identification of a further requirement for free will besides AP, one that also could not be met in a determined world. This is the ultimate responsibility (UR) requirement described when introducing Kane's libertarianism. To recap, here is Kane's account of what the UR entails:

> If agents must be responsible to some degree for anything that is a sufficient cause or motive for their actions (as UR requires), then an impossible infinite regress of past actions would be required unless some actions in the agent's life history (SFAs) did not have either sufficient causes or motives and hence were undetermined.[36]

[35] Unacceptable, that is, not just for the critic of libertarianism but also for libertarians themselves, since if AP are required for free will then it cannot be considered a mere matter of luck which AP a person should in fact choose.
[36] Fischer *et al.* (2007), p. 15.

Kane is concerned not only with our ability to choose between different paths in the future (which the AP requirement seeks to capture), but is concerned also to secure the result that we are the ultimate sources or originators of our acts (which the UR requirement expresses). Since, if determinism were true it could not be the case that one were the ultimate source or originator of one's acts (because one could not be said to be the ultimate source of one's acts if those acts follow inescapably from the state of the world at some time before our birth together with the laws of nature), indeterminism must be embraced. By this means only can we secure the vital condition of UR.

Kane is well aware of the charges levelled at those libertarians who, like himself, wish to make intelligible the UR condition that he describes and endorses. Kane quotes Nietzsche, who once wrote disparagingly that the notion that one could be the ultimate source of one's will and actions is "the best self-contradiction that has been conceived so far," by way of illustrating the fact that there are those who consider the UR condition to be an incoherent and impossible ideal.[37] And in devil's advocate mode, Kane himself poses the question the libertarian must face: "how could acts having neither sufficient causes nor motives be free and responsible actions?"[38]

Kane is thus aware of and able to openly acknowledge the difficulty in rendering the UR condition intelligible; but he nevertheless considers the task possible. His defence of UR relies on a successful defence of his account of SFAs, since not only are occasions when SFAs occur ones at which (according to Kane) we have AP, but they are also a necessary condition for UR. This is because SFAs are the means by which the following regress can be stopped, a regress Kane believes would otherwise rob us of responsibility: agents must be responsible (according to the UR condition) by virtue of past voluntary actions for whatever is a sufficient cause of their actions. In order to stop the regress, certain actions in an agent's life history must be undetermined (and so lack sufficient causes). These actions are, of course, SFAs.

Kane attempts to make his thesis more plausible by explaining that he does not claim that *all* free acts are undetermined, but rather only those acts which help to shape us

[37] Neitzche (1973), p. 32.
[38] Fischer *et al.* (2007), p. 23.

into the kinds of people that we are, namely "will-setting" or "self-forming actions" (SFAs).[39] Kane states:

> I believe these undetermined self-forming actions or SFAs occur at those difficult times of life when we are torn between competing visions of what we should do or become. Perhaps we are torn between doing the moral thing or acting from ambition, or between powerful present desires and long-term goals, or we are faced with competing tasks for which we have aversions. In all such cases, we are faced with competing motivations and have to make an effort to overcome temptation to do something else we also strongly want.[40]

In such situations, Kane suggests that there might be a "stirring up of chaos" within the brain that makes it sensitive to indeterminacies at the neuronal level. The feeling of uncertainty and inner tension that we commonly experience when faced with conflicting sets of motives, Kane postulates, reflects the indeterminacy going on within the cells of our brain.

Let us now recall the argument against incompatibilism that was adopted and reformulated from Ginet:

(1) Incompatibilism entails that (at least) some free actions are not determined by an antecedent state of the world.
(2) If an action is not determined by an antecedent state of the world, then it has (at best) a partial explanation in terms of its antecedents.
(3) All free actions have complete explanations in terms of their antecedents. *Therefore* incompatibilism is wrong.

We saw how Ginet denied premise (3) of the argument, and in so doing rejected the *PSR* which the premise invokes. We also saw how Ginet was able to provide little in the way of evidence to support this denial of the *PSR*, citing unconvincing instances from his own life to demonstrate both causeless action (such as leg-crossing) and also different outcomes following on from the same antecedents (for example, choosing to work or instead opting to watch football). Kane also denies premise (3). The question is: is his case against the *PSR* any stronger?

---

[39] In contrast to Ginet, of course, who argues that an act of will could *only* be free if, at that particular moment, an agent could have done otherwise than they in fact did.
[40] Fischer *et al*. (2007), p. 26.

I would argue that Kane adds little to strengthen the libertarian's case against the *PSR*. As with Ginet, Kane looks to personal experience to corroborate his libertarian theory. He appeals to introspection when he argues that the inner tension we commonly feel at times when we perform SFAs—times when the outcome feels indeterminate—reflects the actual indeterminacy going on at the neuronal level. But as has already been noted, there is no incompatibility between recognising that we can sometimes feel unsure as to the outcome of our deliberations, and believing that there is no indeterminacy to this process. As such, citing these situations as evidence is inadequate. It also seems *ad hoc* of Kane to suggest, as he appears to do, that SFAs are the sole exception to the otherwise exceptionless rule of cause and effect in our universe. Other human acts are presumably determined, as are all acts by other species of animal. So why single out human acts in this way?

None of these considerations against Kane's and Ginet's attempts to persuade us that some actions lack a complete explanation in terms of their antecedents allow us to conclude definitively in favour of the *PSR* and against libertarianism. As I have acknowledged, human action is too complex a phenomenon to yield itself to our understanding in the way that (for example) physical laws do: it is simply not possible either to prove or disprove the claim that, given identical antecedents both within and outside of them, the agent would make the same choice every time. Additionally, the libertarian view has undoubted intuitive appeal precisely because it does not make us beholden to antecedent causes. As Steward urges, part of the appeal of libertarianism lies in its insistence that "it is not because something *about* us makes us act or because something explains why we act, but simply because *we act* that it is up to us what happens to our bodies."[41] Despite these considerations, though, there are better reasons for finding in favour of the *PSR* than there are for rejecting it. While neither proof nor disproof can be provided for the claim that human actions have sufficient reasons, Kane's and Ginet's attempts at providing proof are flimsy. Further, while libertarianism certainly has intuitive appeal, I would argue that the *PSR*'s intuitive appeal is greater still.

We can summarise the 'lack of explanation objection' as follows. Kane, in devil's advocate mode, posed the question: "how could acts having neither sufficient causes

[41] Steward (2012), p. 169.

nor motives be free and responsible actions?" While this question does indeed need answering (which, I have argued, Kane fails to do satisfactorily), a more searching question is simply: "could *any* act lack sufficient causes or motives?" If we accept the *PSR*, then clearly the answer to this is "no." Therefore, the next question to ask must be: does Kane provide good reasons to believe that the *PSR* is mistaken? Once again, I think Kane fails to offer convincing reasons for believing that some actions (namely SFAs) lack sufficient causes or motives. In summary, rather than the *PSR* finding its rebuttal in libertarian theories of will, it is instead libertarian theories of free will that find their rebuttal in the *PSR*.

3.5 Conclusion

I have argued that both Ginet and Kane fail to provide a satisfactory answer to either the 'luck objection' or the 'lack of explanation objection.' I believe we are now in a clearer position to see that it is not some feature peculiar to Ginet's and Kane's own libertarian theories that renders them vulnerable to these objections, but rather that all libertarian theories are equally vulnerable.

The 'luck objection' is applicable to all libertarian theories since all posit indeterminacy in the realm of human action, and as such all are open to the charge that there is no way of explaining this indeterminacy without recourse to the notion of luck. Kane in particular seems to be less than clear about why it is necessary to posit indeterminacy. It is clear enough that there is a connection between indeterminacy and both AP and UR—namely, that indeterminacy is a requirement for both. It is clear too that if you have the intuitive sense that the AP and UR conditions are required for moral responsibility (which is how Kane characterises his own feelings on the matter), you must therefore embrace indeterminacy in order to save moral responsibility. So indeterminacy helps in the sense that it allows for conditions (AP and UR) that Kane intuitively feels are necessary for moral responsibility. But when it comes to explaining the direct connection between indeterminacy and moral responsibility, Kane has nothing convincing to offer. In fact, at times Kane seems to be saying that it is in spite of, rather than because of,

indeterminacy that we are morally responsible agents.[42] In which case, one wonders, what then is the point of being a libertarian?

The 'lack of explanation objection' is likewise applicable to all libertarian theories. The objection is simply that, since the libertarian must affirm the truth of indeterminism, and since the *PSR* commits one to the truth of determinism, the truth of the *PSR* entails the falsity of libertarianism. Relying as it does on the truth of the *PSR*, this 'lack of explanation objection' is perhaps a less sure route to undermining libertarianism than the 'luck objection.' Indeed, Ginet and Kane both see no problem in rejecting the *PSR*, the former citing examples of uncaused action (such as leg-crossing) and of different choices following on from identical antecedent conditions as evidence of its falsity, and the latter appealing to our supposed intuition that, when faced with difficult decisions that create inner tension, the outcome is indeterminate.

In both cases, however, I called into question the strength of the evidence against the *PSR* and hence against the 'lack of explanation objection.' In particular, Ginet's examples of *PSR*-violating actions and choices strike me as wholly inadequate since the actions and choices he picks out could just as well be considered determined, and nothing he says would persuade those not already convinced that some human actions are undetermined. Kane also does little to persuade us to reject the *PSR*, beyond appealing to the common experience of feeling as though, when making certain difficult decisions, the outcome is not determined. While we may or may not agree with Kane that we sometimes sense this absence of determinacy in the outcome of our decisions, this remains a poor basis on which to conclude that indeterminacy is therefore present. Still, to Kane's credit he does offer a fuller explanation than Ginet of how indeterminacy might enter the picture, arguing that indeterminacy at the quantum level could have macro level effects on account of the stirring up of chaos within the brain.[43] The overall impression, though, is that while

[42] In the following passage, for example, Kane (in Fischer *et al.* (2007: p. 27)) suggests that the process of solving a mathematical puzzle is analogous to making the correct moral decision since both, he claims, are feats accomplished despite the hindrance of indeterminacy: "Whether you are going to succeed in solving the problem is uncertain and undetermined because of the distracting neural noise. Yet, if you concentrate and solve the problem nonetheless, we have reason to say you did it and are responsible for it, even though it was undetermined whether you would succeed. The indeterministic noise would have been an obstacle that you overcame by your effort."

[43] Fischer *et al.* (2007), pp. 28-31.

Kane's explanation does at least attempt to marry libertarianism with a scientific understanding of the world, his appeal to chaos theory and to the results of standard quantum theory in order to explain indeterminacy in human action is unacceptably *ad hoc*. In brief, while Kane's explanation is not demonstrably false, there is little reason to consider it true.

In conclusion, I consider the 'luck objection' on its own to provide sufficient reason to reject libertarianism in all of its forms. Given that it successfully shows that indeterminacy cannot allow for any more freedom or moral responsibility that we would otherwise lack in a determined world, I cannot see any reason to adopt the libertarian position. However, the 'luck objection' does not rule out the possibility of indeterminacy in human actions *simpliciter*, but rather demonstrates that indeterminacy can be of no use in defending our human freedom and/or moral responsibility against the perceived threat of determinism. The 'lack of explanation objection,' by contrast, *does* rule out the possibility of indeterminism and thus, if successful, not only leaves one with no reason to adopt a libertarian position but positively precludes its adoption. Admittedly, Ginet, Kane, and their fellow libertarians are not compelled to agree that the 'lack of explanation objection' is successful—indeed, their *PSR*-defying accounts of human action testify to their disagreement with this line of argument. But while libertarians can deny the *PSR* in this way without inviting accusations of contradiction or absurdity, there is little reason to accept their accounts of human action unless one has prior grounds for endorsing libertarianism. With the connection between indeterminism and increased freedom and/or moral responsibility severed by the 'luck objection,' no prior grounds for endorsing libertarianism can be found, and thus it looks as if libertarians are requiring us to reject a principle for which we have a presumption in favour while getting nothing in return.

4

The Case for Hard Determinism

We begin this chapter having established, if not conclusively then at least to our satisfaction, that libertarianism is false, and the *PSR* - and hence also determinism—is true. As has been shown, the presumed truth of the *PSR* entails the truth of determinism, and neither the arguments against the *PSR* as presented in Chapter 2 nor our examination of the case for libertarianism in Chapter 3 can persuade us to presume otherwise.

The central question to address now is: what implications does the truth of determinism have for our ascriptions of moral responsibility?[1] In particular, we shall be asking whether there are good arguments for thinking that the truth of determinism precludes the possibility of holding each other morally responsible for our acts. Arguments to this effect, if successful, would lead us to adopt an 'error theory' regarding our ascriptions of moral responsibility: that is, we should be forced to conclude that holding people morally responsible for their actions is always a mistake, and we can only continue to do so on pain of illogicality.

At this stage, it might be useful to present a table to illustrate the various permutations of opinion about determinism and freedom that it is possible to have, whether one holds to their truth (T), their falsity (F), or else remains agnostic (?):[2]

[1] I could equally have asked what implications the truth of determinism has for our ascriptions of *free will* to one another—that is, I take it that when we discuss whether or to what extent we possess free will, we are talking about the *free will required for moral responsibility*. So on my understanding (or, perhaps, stipulation) free will is always and only present in actions for which we are morally responsible, and *vice versa*.

[2] An identical table to the one presented here can be found in Pereboom (2001: xix), and also in Strawson (2010: p. 6).

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Determinism	T	F	T	F	T	F	?	?	?
Freedom	F	T	T	F	?	?	F	T	?

As a result of our investigations, we can already rule out positions (2), (4), (6), (7), (8), and (9). We are left with a choice of affirming either positions (1), (3), or (5)—determinism is true, and we can either affirm that we have the free will required for moral responsibility, deny this, or else (for whatever reason) remain agnostic about the possibility of freedom in the face of determinism.

There are two arguments to examine in this chapter in favour of position (1), the incompatibilist position that we lack the free will required for moral responsibility.[3] Each of these arguments will be presented and then followed by objections to see whether their conclusion—essentially, that freedom and determinism are incompatible—can be avoided. I have chosen these arguments on the basis that they seem to me to be the most compelling arguments for the incompatibilist position. Any candidate theory of moral responsibility, therefore, will have to provide a convincing response to these arguments if it is to be credible.

4.1 The Consequence Argument

The first argument to consider is known as the Consequence Argument. It has received various presentations at the hands of different philosophers, but the version

[3] In actual fact the second argument, which is proposed by Galen Strawson, might be more accurately classified as an argument for position (7), since Galen Strawson does not affirm the truth of determinism but instead argues that, whether or not determinism is true, the free will required for moral responsibility that we are generally assumed to possess is an impossibility. Nevertheless, I will be taking Galen Strawson's argument as providing a defence of (1), since unlike Galen Strawson I *do* see compelling reasons to affirm the thesis of determinism, at least as far as human actions are concerned.

that we shall focus on is expounded by van Inwagen.[4] Here is how the argument runs:

(1) No one has power over the facts of the past and the laws of nature.
(2) No one has power over the fact that the facts of the past and the laws of nature entail every fact of the future, including one's own actions.

Therefore no one has power over the facts of the future, including one's own actions.[5]

The argument is admirably straightforward, both in its premises and its conclusion. The basic idea is that powerlessness over both facts about the past and laws of nature transfers to powerlessness over one's own actions, as these actions are the inescapable consequences of facts about the past combined with the laws of nature. If free action requires the ability to do otherwise than one actually does, and if determinism is true, then it seems as though it is never possible to do otherwise than what one actually does, and hence that it is never possible to act freely.

It should be noted that while van Inwagen accepts the conclusion of the argument, he is not himself a hard determinist since he does not accept the truth of determinism. For him, then, the argument functions as a *reductio* of the affirmation of determinism—since free will is impossible if determinism is true, we can safely reject the thesis of determinism. Of course, it is not possible to view the argument in this way if, as I have argued is the case, there is good reason to think that determinism is true. For those of us that reject the libertarian metaphysics of van Inwagen and others, the argument poses a serious threat to the existence of free will.

There are two lines of response to this argument to consider. The first is from Lewis, who seeks to exploit a perceived equivocation in the Consequence Argument in order to avoid the conclusion. The second response comes from Slote, who questions the legitimacy of moving from the notion of being powerless over facts that are in no way alterable by our beliefs, desires, and abilities etc., to the notion that one is equally powerless over facts that, conversely, are the result of what Slote calls

[4] Van Inwagen first presents the Consequence Argument in his (1975) paper "The Incompatibility of Free Will and Determinism." Other presentations of the argument have been offered by the following: Ginet (1966); Lamb (1977); and Wiggins (1973). While there are minor variations in emphasis between the different presentations, all agree that, if all human actions are in some sense necessary, then we lack the free will and/or moral responsibility we commonly attribute to ourselves.
[5] Van Inwagen (1983), p. 222.

"appropriate internal factors."[6] After presenting and critiquing these responses, we shall conclude by pronouncing on the success (or otherwise) of the Consequence Argument.

4.2 Lewis's Critique

Defining himself as a compatibilist ("the doctrine that soft determinism may be true"), Lewis says that he shall feign to uphold soft determinism ("the doctrine that sometimes one freely does what one is predetermined to do") for the sake of argument.[7] He then proceeds to outline the following scenario:

> I have just put my hand down on my desk. That, let me claim, was a free but predetermined act. I was able to act otherwise, for instance to raise my hand. But there is a true historical proposition *H* about the intrinsic state of the world long ago, and there is a true proposition *L* specifying the laws of nature that govern our world, such that *H* and *L* jointly determine [that I put my hand down on the desk].[8]

Lewis then asks: what would have had to have been the case if, instead of putting his hand on the desk, he had at that moment raised it? The answer is that at least one of three things would of necessity been true. Either contradictions would have been true together; or the historical proposition *H* would not have been true; or the law proposition *L* would not have been true.

Not surprisingly, this admission amounts to a *reductio* for van Inwagen, as none of the three alternatives at first glance seem acceptable. Lewis himself discounts the possibility that if he had raised his hand then contradictions would have been true together, as well as the possibility that raising his hand would have altered H, the historical proposition about the intrinsic state of the world long ago. But he is happy to be committed to the consequence that if he had raised his hand as he claims he was able to do, then some law would have been broken and *L* would thus have been rendered false.

A first thought at hearing this might be to wonder whether it is not part of the very definition of a law that it cannot be broken, in which case it is a conceptual necessity

---

[6] Slote (1982), p. 21.
[7] Lewis (1981), p. 122.
[8] *Ibid.*, p. 122.

that the law proposition *L* must remain true. Lewis is aware of this, however, and does not wish to claim that anything could be both a law and be broken, since on his understanding a genuine law must be an absolutely unbroken regularity. Instead, he clarifies, his contention is: "If *L* had not been true, something that is in fact a law, and unbroken, would have been broken, and no law."[9] Notwithstanding this clarification, it might reasonably be wondered how it could be the case that we are able to perform acts that would entail the falsity of the law proposition *L*. Lewis imagines his philosophical adversary crowing: "You claim to be able to break the very laws of nature. And with so little effort! A marvellous power indeed! Can you also bend spoons?"[10] However, Lewis goes on to argue that his detractor misunderstands the nature of his claim. What his philosophical adversary is actually objecting to is something that Lewis terms the 'Strong Thesis,' which states:

ST: I am able to break a law.

Lewis, though, stresses he is in complete agreement that ST is utterly incredible, and as such he does not wish to make a plea for its truth. What Lewis does wish to argue the case for is the following 'Weak Thesis':

WT: I am able to do something such that, if I did it, a law would be broken.

Lewis proceeds to defend WT by sketching a scenario in which he is able to raise his hand "although it is predetermined that I will not."[11] Lewis argues that all that is required for him to raise his hand in a situation in which it is predetermined that he will not in fact do so, is for some law to be broken. Crucially for Lewis, however, his act of hand-raising, were it to happen, would itself neither cause nor be a law-breaking event. Neither would it be the case that *any* act of his would either cause or be a law-breaking event. Therefore, his alleged ability to raise his hand confers upon him no "marvellous ability" to break laws of nature. Instead, all that need be the case is that before the act of hand raising a law must have been broken, an event that Lewis terms a "divergence miracle."

At this point, it would be helpful to pause and consider how Lewis's embracing of WT relates to van Inwagen's Consequence Argument. Lewis is arguing that it is

[9] *Ibid.*, p. 123.
[10] *Ibid.*, p. 123.
[11] *Ibid.*, p. 124.

possible to perform certain actions (such as raising one's hand) whose performance would render false proposition *L* (which specifies the laws of nature governing our world). As such, the accusation levelled at van Inwagen is that he is guilty of equivocation between the following two propositions, either of which could be said to be entailed by premise (1) of the Consequence Argument:

(PROP 1) I am unable to do anything such that my act would be or cause a law-breaking event.

(PROP 2) I am unable to do anything such that were I to do it, an actual law of nature would be false (and hence not a law).

Lewis concedes that (PROP 1)—which is effectively the denial of the truth of ST stated above – is impossible; but he denies that (PROP 2)—which is the negation of WT—is likewise impossible. According to Lewis, the Consequence Argument is now either invalid (if we are to understand premise (1) as entailing the true (PROP 1)) or else it is unsound (if premise (1) is instead meant to entail the false (PROP 2)).

4.2.1 Response to Lewis

Does Lewis's attempted rebuttal of the Consequence Argument stand up to scrutiny? I would argue not, and that detailed consideration of his defence of compatibilism will reveal this fact. A first point to note is that, while Lewis identifies an equivocation in van Inwagen's argument—premise (1) can either be interpreted as entailing (PROP 1) or (PROP 2)—van Inwagen would not consider himself guilty of equivocation since his premise is intended to affirm the truth of both. That this is so is clear from a later article in which van Inwagen clarifies what he means when he claims that no one is able to render propositions stating the laws of nature false:

> An agent was able to render a proposition false if and only if he was able to arrange things in a certain way, such that his doing so, together with the whole truth about the past, strictly implies the falsity of the proposition.[12]

This definition makes it plain that van Inwagen would accept neither ST nor WT, and so would endorse both (PROP 1) and (PROP 2) since these propositions are the

[12] Van Inwagen (2004), p. 346.

negations of those theses. Nevertheless, that van Inwagen is not guilty of equivocation is not to say that Lewis is wrong to claim that there is a distinction between (PROP 1) and (PROP 2); neither does it show he is wrong to think it possible to endorse one without endorsing the other; nor yet does it show that his contention that (PROP 1) is true while (PROP 2) is false is mistaken. In fact, we certainly can make a distinction between (PROP 1) and (PROP 2), and Lewis is perfectly entitled to argue for the truth of the former and the falsity of the latter.

However, there are good reasons to think that Lewis is unsuccessful in urging the falsity of (PROP 2), and thus that his challenge to the Consequence Argument is not successful either. First, we can call into question whether it is any more plausible that we should possess weak abilities than that we should possess the allegedly much more implausible strong abilities. Recall that weak abilities are supposed to be more acceptable and less open to ridicule than strong ones on account of the fact that weak abilities, unlike their strong counterparts, do not confer upon us the "marvellous ability" to break laws of nature. Evidently, this fact is highlighted by Lewis in order to underline the point that we should not balk at his endorsement of weak abilities. But is what Lewis proposes any easier to accept than if he were to endorse both weak *and* strong abilities? Both weak and strong abilities, it can be argued, are equally implausible (or, even less charitably, equally impossible) since *both require that a law of nature be broken*. So it is no defence for Lewis to argue that his compatibilism is plausible since he does not accept ST: his acceptation of WT alone is enough to commit him to a belief in the possibility of miracles, and hence his compatibilism remains suspect.[13]

Lewis certainly does not consider the theses WT and ST to be equally plausible, as the examples he uses to illustrate the two theses reveal. It will be recalled that, in defending WT, Lewis appeals to his ability to raise his hand even though it is predetermined that he will not do so. It is Lewis's supposed possession of this ability that demonstrates the truth of WT, which asserts: I am able to do something such that, if I did it, a law would be broken. No "marvellous power" is required of him in order for this assertion to be true—the raising of one's hand is, after all, a rather mundane act. What is needed (or what is *merely* needed, Lewis might say), is for his

[13] In fact, it has been argued that Lewis's theory commits him to upholding not just WT but also ST (see Beebee (2003: p. 272)). I will not press this objection, however.

act of hand-raising to be preceded, whether directly or at some time in the distant past, by a "divergence miracle" that falsifies a law. By contrast, Lewis implies (and presumably believes) that marvellous abilities *are* required for ST to be true, and this belief helps to persuade him that it is not. As an example of an act that would *cause* a law-breaking event, Lewis suggests the following: "Suppose that I were able to throw a stone very, very hard. And suppose that if I did, the stone would fly faster than light, an event contrary to law. [In such an event] I would be able to do something such that, if I did it, my act would cause a law-breaking event."[14] And for an example of an act that is *itself* a law-breaking event, Lewis imagines: "Suppose that I were able to throw a stone so hard that in the course of the throw my own hand would move faster than light. [In such an event] I would be able to do something such that, if I did it, my act would itself be a law-breaking event."[15]

There is nothing wrong with the above examples *per se*: the case of raising one's hand is an example of a weak ability in action, and the examples in which a stone is thrown variously either *cause* a law-breaking event or is *itself* a law-breaking event are indeed valid examples of strong abilities. The problem is that the examples imply, falsely, that strong abilities necessarily require some manifestly miraculous power on the part of the agent, a power that it is clearly implausible to suppose that we possess. Set against strong abilities thus presented, weak abilities are thereby made to seem all the more plausible—a weak ability, after all, appears to be identical to many everyday acts such as raising one's hand.

But in actual fact the distinction between weak and strong abilities has nothing to do with the size of miracle required for their possession. Rather, the distinction is simply to do with whether or not the miracle is (or is caused by) an act of mine.[16] To see that a strong ability need not appear all that miraculous, we can construct our own mundane example. In fact, Lewis's earlier example of hand-raising can be appropriated for this purpose: all that needs to be done is to stipulate that the act of hand-raising takes place not because of some earlier divergence miracle, but rather that the hand-raising is *itself* the divergence miracle. And so now we have imagined two identical-looking hand-raising scenarios: the first involves a weak ability and is

[14] Lewis (1981), p. 124.
[15] *Ibid.*, p. 124.
[16] Beebee (2003: p. 272) makes the same point about Lewis misrepresenting the distinction between strong and weak abilities.

made possible by the occurrence of an earlier divergence miracle; while the second involves the possession of a strong ability and, though the event appears unassuming, it is in fact a miracle. The question for Lewis is: why think that the former scenario is any more plausible than the latter? After all, both require a miracle, and neither is any more or less miraculous than the other. There seems, therefore, to be no principled reason for assuming that no divergence miracle could ever be a human act. Such an act would be no more miraculous (and therefore no less plausible) than an event that does not involve an agent.

In conclusion, van Inwagen and his fellow incompatibilist need not and should not accept either WT or ST. Both require for their truth the occurrence of a miracle and the consequent breaking of what would otherwise be a law of nature, and we have seen already how the implication that the truth of ST would require a more miraculous miracle (so to speak) than the truth of WT is misleading. All that van Inwagen needs to say – and what he does in fact say—is that the supposed ability to act in such a way that, if one were to perform that act then the proposition *L* expressing the laws of nature would be falsified, is not an ability we ever possess.[17]

But it is not just the case that the truth of WT is implausible, or that we have little reason to accept it (although these are both true enough). More than that, it surely *cannot* be true if we accept the thesis of determinism on the basis that it is entailed by the truth of the *PSR*. To see this, consider how Beebee elucidates Lewis's notion of a divergence miracle: "[D]ivergence miracles are themselves events, and, by definition, the closest world where a divergence miracle *f* occurs is a world whose facts do not diverge from those of the actual world until *f* itself occurs."[18] This passage makes clear that a divergence miracle is an occurrence that marks a bifurcation between two previously identical worlds: that is, two worlds sharing identical histories diverge on the occasion of a divergence miracle in one or other of those worlds. Considering this explanation in the light of our acceptance of the *PSR*,

[17] Van Inwagen (2004: p. 348) speaks in terms of Lewis and his fellow compatibilist having to "measure the price" in response to the Consequence Argument, a phrase that Lewis himself has used when writing about philosophical argumentation. In this context, van Inwagen says, the price is that the compatibilist "must believe that a free agent in a deterministic world is able to arrange things in such a way that one's so arranging them, together with the whole truth about the past, strictly implies the falsity of at least one law of nature," a situation that he describes as "obviously impossible."
[18] Beebee (2003), p. 269.

it is evident that no such divergence between hitherto identical worlds could take place, since there could be no explanation for this occurrence.[19]

4.3 Slote's Critique

We have examined Lewis's critique of the Consequence Argument and found it wanting. Let us now examine a different critique, from Michael Slote, who presents an argument to the effect that our powerlessness over the past and the laws of nature to which the Consequence Argument makes reference (and which he accepts is the case) depends on a selectivity that does not apply to present and future acts. That is, we are powerless to alter either the past or laws of nature because current desires and abilities are ineffective in bringing about changes to facts about these two things. By contrast, current desires and abilities are indisputably causally relevant to present and future actions. So in the sense that future actions are at least partly the result of current beliefs, desires, abilities and so forth, it is possible (contrary to what the Consequence Argument would have us believe) to alter future actions.

This rough sketch of Slote's response to the Consequence Argument will now be fleshed out before we consider its merits. To begin with, let us present the basic argument form that Slote believes underlies all versions of the Consequence Argument, and which he will therefore be criticising: [20]

[19] I have presented here a case for rejecting WT if one accepts determinism on the grounds that it is entailed by the *PSR*. However, I think there is a good case for simply saying that WT is incompatible with determinism *tout court*, and so no further appeal to the truth of the *PSR* is necessary. This is effectively van Inwagen's position when he argues that it is obviously impossible for an agent to arrange things so that this arranging, together with the whole truth about the past, strictly implies the falsity of *L*. If this is correct, then Lewis's positing the existence of divergence miracles amounts to him claiming that determinism is false; and if this is the case, then what relevance does his objection have for the Consequence Argument, an argument that assumes the truth of determinism?

[20] Although van Inwagen does not formulate his argument in this fashion, Slote (1982: p. 10) certainly thinks that he indirectly appeals to this form of argumentation, which we see entering his incompatibilist argument "in pieces, rather than whole." I think Slote is right that this formulation of the Consequence Argument is faithful to van Inwagen's intentions, and so it provides a legitimate basis for Slote's critique.

Np	(where 'p' stands for a statement that posits the existence of some earlier event or circumstance)
N ($p \supset q$)	(where '$p \supset q$' stands for some law of nature)
_____	
∴Nq	(where 'q' stands for a statement that posits some human action)

Slote adds that the necessity indicated by 'N' is not to be considered a form of logical necessity. Instead, 'Np' can be thought to abbreviate 'p, and no one has, or ever had, any choice about whether p.'

After stating the argument, which Slote terms the "main modal principle," he next seeks to demonstrate that the sort of inference found here from Np and N ($p \supset q$) to Nq is questionable.[21] It is questionable because it assumes that the necessity expressed in the operator 'N' is: (a) agglomerative, and; (b) closed under entailment. Speaking generally, agglomeration involves conjoining two propositions to make a third, in the belief that if the former two are true then the latter will also be true. In this particular context, the assumption of agglomeration amounts to assuming the validity of moving from 'Np' and 'N ($p \supset q$)' to 'N (p. $p \supset q$)'. Meanwhile, the assumption that the operator 'N' is closed under entailment sanctions the move from 'N (p. $p \supset q$)' to the conclusion of the argument, which is 'Nq'.

Slote's next aim is to demonstrate how, under certain circumstances, the assumptions of agglomeration and of closure under entailment are not warranted. Beginning with a fairly uncontroversial example, Slote notes that it is generally accepted that 'A knows that p' and 'A knows ($p \supset q$)' do not entail 'A knows that q.' This lack of entailment is, of course, because people sometimes fail to make inferences that they

[21] While Slote rejects this "main modal principle" as embraced in one form or another by those who champion the Consequence Argument, he compares it favourably to what he calls the old form of deterministic argument against freedom of will, which has roughly the form: p, N ($p \supset q$) ⊢ Nq. This old form is clearly fallacious, Slote asserts, as if no necessity attaches to past or pre-existent events posited by p, then neither will there be any necessity to the human action posited by q.

are entitled to make. If we accept this example, then we accept that closure under entailment sometimes fails.

As Slote acknowledges, however, this example is of limited use since closure under entailment fails here due to epistemic weakness on the part of *A*, whereas the success of the entailment relation in the Consequence Argument is not dependent on anyone's ability to draw the correct inference. Still, there are further examples, Slote insists, that show we should not take agglomeration and closure under entailment for granted, with one other category of examples being those involving obligations. We are asked to consider a situation in which a promise has been made to a friend to return their book, and a further promise has been made to another friend to meet them at a certain time. Slote suggests that, on making these promises, it will be obligatory that one returns the book, and obligatory that one meets the friend; but not obligatory that one performs the joint act of returning-the-book-and-meeting-one's-friend. This example demonstrates that obligations are not agglomerative, argues Slote, and they are not agglomerative because they involve relations between specific people. Obligations can also fail to be closed under entailment, Slote informs us, citing as an example the fact that being under an obligation to meet a person at three o'clock tomorrow does not entail being under an obligation to stay alive until tomorrow, despite the fact that staying alive is a precondition for meeting.

The champion of the Consequence Argument might well argue that these examples have little to do with the case in hand: while the Consequence Argument is concerned with the rules of alethic modal logic, the first example above concerns epistemic modalities while the second concerns principles of inference from deontic modal logic. But Slote also has an example to hand from the realm of alethic modal logic, one which he believes offers a credible counterexample to the assumption that examples from this realm must obey agglomerativity and closure under entailment. The example is of cases in which events that are themselves not accidental can sometimes come together in order to create an event that is accidental. Imagine, we are asked, that two friends, Jules and Jim, have each been sent to the same location at the same time by their superiors as part of their work (they might, for example, have both been sent to a bank by their respective bosses in order to withdraw money). In such a case, Slote says, "it would appear to be no accident (not accidental) that Jules is at the bank when he is, no accident (not accidental) that Jim is there when he is,

but a benign and lovely accident (accidental) that Jim and Jules should both be there at that time."[22]

The moral that we are to draw from this example is that the non-accidental is not agglomerative:

> Most people who take the trouble to think about it would recognize it to be a main feature of (non)accidentality that things that in themselves appear perfectly regular and nonaccidental may "come together" to create something that is accidental, and this feature precisely is the non-agglomerativity of the accidental.[23]

Not content with wishing to show that the non-accidental is not agglomerative, Slote also claims that the non-accidental appears not to be closed under entailment either. We can imagine that it might be no accident for Slote to be in a certain place at a certain time (having been sent there by a superior in his line of work), yet it might nevertheless be an accident that he remains alive right now (having narrowly escaped getting flattened by a truck a few minutes ago). What this shows, Slote thinks, is the following:

> If closure thus fails for nonaccidentality, we can perhaps go on to deny that it is governed by our main modal principle [i.e. the form of inference on p. 91], on the grounds that [in the above example] where it is no accident that I am at a certain place and yet something of an accident that I still exist, it is also (on trivial logical grounds) no accident that if I am at that place, then I still exist.[24]

Now we are in a position to articulate what it is that Slote thinks is wrong with the Consequence Argument. Just as non-accidentality is not agglomerative or closed under entailment, so too is the notion of unavoidability appealed to in the Consequence Argument not agglomerative or closed under entailment. Further, both non-accidentality and unavoidability fail to be agglomerative and closed under entailment for the same reason, which is that both notions have a selectivity built into them. We have seen how non-accidentality is selective from the above examples, in which this feature or property exists only in relation to a plan of a specific kind that calls for particular events to occur. How then is the notion of unavoidability as employed in the Consequence Argument selective?

---

[22] Slote (1982), p. 15.
[23] *Ibid.*, p. 15.
[24] *Ibid.*, p. 16.

One specific way in which the notion of unavoidability that we find in the Consequence Argument is selective, Slote explains, is that when we say that some event before our birth (which is where the Consequence Argument usually begins) is something we can now do nothing about, we can be interpreted as meaning that the event is beyond the causal reach of our desires, beliefs, and abilities acquired during our lifetime. Speaking generally, "the factor "selected" by such necessity, is, simply, some factor (or set of factors) that brings about the unavoidable thing *without making use of (an explanatory chain that includes) the desires, etc., the agent has around that time.*"[25]

With these points in mind, Slote now has the ammunition to mount an argument against the Consequence Argument. Although past events are beyond an agent's control, as are the laws that lead from them to the agent's actions, "it will not follow that those actions are themselves necessary at some later time when the agent is considering whether to perform them."[26] This is because, although the actions are determined by and therefore (in principle) predictable in terms of factors prior to the agent's desires, abilities and so forth, those earlier factors bring about an agent's actions only by means of causal chains in which the agent's desires, abilities and so forth are involved. Once agential desires, abilities etc. are engaged in the right ways, an agent's actions can no longer be correctly characterised as unavoidable, and therefore the Consequence Argument fails. In summation, Slote writes:

> The selectivity that is plausibly attributed to the necessity involved in recent defences of incompatibilism distinguishes between factors within and factors external to agents. For, on our rough gloss, what is effected without the "help" of an agent's (coeval) character, abilities, desires, etc. is (then) unavoidable, unalterable, inevitable, but what is brought about via a causal chain that includes appropriate internal factors is not."[27]

4.3.1 Response to Slote

Slote believes that an agent is able to render *at least some of* the consequences of the propositions concerning the past and the laws of nature false. His justification for

[25] *Ibid.*, p. 19.
[26] *Ibid.*, p. 20.
[27] *Ibid.*, p. 21.

believing this is that some of these consequences come about (at least in part) due to the agent's character, beliefs, desires, and so forth.

But what force does this objection have against van Inwagen's argument? After all, it is not as though van Inwagen has let the matter of distinguishing between internal and external factors escape unnoticed—indeed, as Slote concedes, van Inwagen denies that such a distinction in any way affects the soundness of the Consequence Argument.[28] Part of the reason Slote believes his objection has force against the Consequence Argument is that he thinks his various examples demonstrate the existence, previously unconsidered, of cases of alethic necessity that are not agglomerative or closed under entailment. Slote writes:

> Previous suggestions that we analyze ability conditionally in order to defend compatibilism from the argument of [Consequence Argument proponents] have seemed ad hoc in the absence of alethic necessity not subject to our main principle or agglomeration under closure; but when we supply such examples and explain how various necessities can flout those principles, the selective necessity that we have claimed may be involved in the free-will controversy is made to seem part of a plausible pattern, rather than an isolated case.[29]

So "conditional analyses" of ability in response to the Consequence Argument were *ad hoc*, Slote believes, in the absence of examples of other sorts of alethic necessity that are non-agglomerative and/or not closed under entailment. Now that these examples have been provided, however, it is no longer *ad hoc* to make the case for some sort of conditional analysis.

It is true that Slote provides some ingenious examples to support his case. However, I would question whether the examples he offers are really sufficiently similar to the Consequence Argument situation to provide much support for what is, when all is said and done, quite a straightforward objection (and one, moreover, that van Inwagen and other incompatibilists have considered and rejected). For instance, although Slote characterises the cases in which he discusses non-accidentality as examples of alethic modal logic, this characterisation is questionable. It is questionable because the judgment regarding whether some event is or is not an accident surely depends on how much we know, and it is this fact that explains why

[28] See van Inwagen (1975: pp. 196-7) for his discussion of the so-called "conditional analysis" of ability, and his claim that the validity of his Consequence Argument in no way depends upon demonstrating the inadequacy of this typically compatibilist analysis of ability.
[29] Slote (1982), p. 22.

two non-accidental events can combine to create what we judge to be an accidental event. To take Slote's example as recounted earlier, it is a consequence of Jules and Jim's incomplete knowledge of each other's movements that makes the non-accident of Jules being sent to the bank, and the non-accident of Jim being sent to the bank, combine to create the benign accident of them both meeting at the bank. If each of them knew more about the other's work and the kinds of errands that the other is expected to run, then their meeting would no longer be accidental since their increased knowledge of the other's movements would lead each of them to anticipate the encounter.

There is no such parallel in the case of unavoidability, no comparable way in which the notion of the unavoidable can be said to escape the rules of alethic modal logic. As such, there is a pertinent difference between the two cases, and this should make us question whether the non-agglomerativity and absence of closure under entailment in the case of the non-accidental really puts any pressure on the Consequence Argument proponent to think that unavoidability might be similarly non-agglomerative and not closed under entailment.

We have reached here a familiar impasse between, on the one hand, compatibilists who have the intuition that it is possible to do otherwise in a determined world and, on the other hand, incompatibilists who deny this possibility. In the clash of intuitions here, I favour the incompatibilist position: it seems to me that the soundness of the Consequence Argument is undeniable. Though their arguments differ, Slote's and Lewis's objections both rely for their persuasiveness on the prior adoption of a compatibilist understanding of ability. Lacking this compatibilist perspective on the notion of ability, I fail to be persuaded.

4.4 The Impossibility of Moral Responsibility Argument

We have examined van Inwagen's argument to the effect that, if determinism is true then we lack the free will required for moral responsibility. However, we turn now to an argument that makes the stronger claim that, regardless of the truth or falsity of

determinism, it is impossible for us to possess the free will required for moral responsibility.[30] The argument is due to Galen Strawson, and is titled the Impossibility of Moral Responsibility Argument. It runs as follows:

(1) When you act, you do what you do—in the situation in which you find yourself—because of the way you are.
(2) If you do what you do because of the way you are, then in order to be ultimately morally responsible for what you do you must be ultimately responsible for the way you are.
(3) You cannot be ultimately morally responsible for the way you are.
Therefore you cannot be ultimately morally responsible for what you do.[31]

The first premise claims, pretty uncontroversially, that an agent's action in any given situation is the product of the way that they are—that is, it is a product of their character. To take a simple example, a person who is temperamentally disposed towards anger might find themselves beeping their horn extremely loudly when stuck in traffic on a hot summer's day. By contrast, a person with an altruistic disposition might in the same situation find themselves reflecting with sadness on the plights of the many other unfortunate souls behind them in the traffic jam whose sweltering ordeal will last even longer than their own. According to the first premise, then, how we act is a function of the way we are, a function of our character. As I have said, it is not a controversial claim, and its truth is not at issue in any of the critiques of the argument that we will be considering presently.

The second premise takes the first premise as an antecedent (the claim that you do what you do because of your character), and claims in the consequent that to be ultimately morally responsible for your actions you must be ultimately morally responsible for your character also. The thought here is as follows: since your actions derive from your character, it must be the case that ultimate moral responsibility for one's actions entails ultimate moral responsibility for one's character. One question that this premise raises is just what it means to be *ultimately* morally responsible for something, whether for one's acts, or character, or for something else entirely. How

[30] It is not strictly speaking a hard determinist argument, therefore, since the truth of determinism features neither as a premise, nor as an assumption underlying any premise, of the argument. Since we have prior reasons to accept determinism, however, we can view the argument as providing grounds to accept hard determinism if it is successful.
[31] Strawson (2002), p. 443.

does mere *moral* responsibility different from *ultimate* moral responsibility? Strawson's explanation runs as follows:

> [Ultimate moral responsibility] is responsibility and desert of such a kind that it can exist if and only if punishment and reward can be fair or just without having any pragmatic justification, or indeed any justification that appeals to the notion of distributive justice.[32]

So *ultimate* moral responsibility as opposed to mere moral responsibility, thinks Strawson, can have nothing to do with pragmatic justifications. To take an example, while some might argue that at least part of the purpose of holding people morally responsible for their acts is to shame them into behaving differently in the future, this sort of understanding of moral responsibility is incompatible with the notion of *ultimate* moral responsibility that Strawson presents in the second premise of his argument. Similarly, consigning a dangerous criminal to a lengthy stay in prison in order to prevent them from committing further heinous crimes may legitimately be seen as a way of holding that criminal morally responsible for their behaviour; but since the rationale for their sentencing is in this scenario purely pragmatic (i.e. they are incarcerated in order to prevent further crimes being committed) this punishment cannot be seen as a judgment on whether the criminal is *ultimately* morally responsible for what they have done.

The third and final premise is the shortest—yet for many libertarians the most contentious—of the premises in Strawson's argument. This premise states that no-one can be ultimately morally responsible for the way they are—and, since being ultimately responsible for one's character is a precondition for being ultimately morally responsible for one's acts, the conclusion follows that it is impossible for anyone to be ultimately morally responsible for anything that they do.

Why are we not—Indeed, why is it not so much as *possible* for us to be—ultimately morally responsible for the way we are (that is to say, for our character)? The answer is that ultimate moral responsibility for your character is impossible as this would require the completion of an infinite series of acts of self-formation. To see this, we can begin by observing that, in order to be ultimately responsible for your present character, it would have to be true that you intentionally brought it about at some earlier time that you have the character you currently possess. But of course, in order

[32] *Ibid.*, p. 452.

for this to be the case, prior to this act of character formation you must have had a certain mental nature on account of which you acted to bring it about that you have the character that you now do. And this mental nature must have been intentionally formed due to some prior nature on account of which you acted, and so on *ad infinitum*.[33]

This is not to say that Strawson denies that we commonly feel ourselves to be ultimately responsible for our characters—on the contrary, he acknowledges that we often unreflectively (and incorrectly) experience ourselves as ultimately responsible, in just such a way that would only be warranted if the above process of self-formation or self-determination could actually be completed.[34] Strawson is therefore urging us to abandon this unreflective tendency both to assume ultimate responsibility for our own character and also to ascribe ultimate responsibility to others for theirs. Nietzsche is mentioned approvingly in this connection for having declaimed: "*No one* is accountable for existing at all, or for being constituted as he is, or for living in the circumstances and surroundings in which he lives."[35] Strawson's argument effectively adds that, as a consequence, no one is ultimately accountable for anything whatsoever. Consequently, claims Strawson, there is a "fundamental sense in which no punishment or reward is ever ultimately just."[36]

4.5 Hurley's Critique

On account of the arguments of the previous two chapters, libertarianism has now been ruled out of the picture. This makes our task in this section somewhat easier, since libertarians are inclined to deny either premise (1) or (3), while compatibilists typically direct their attention to demonstrating the falsity of premise (2).[37] As such,

[33] *Ibid.*, pp. 445-7. The irresoluble dilemma for Strawson's opponent, as Strawson himself articulates, is that "[t]here has to be, but there cannot be, a starting point in the series of acts or processes of bringing it about that one is a certain way, or has a certain nature, a starting point that constitutes an act or process of ultimate self-origination."
[34] See Strawson (1986), p. 106.
[35] Nietzche (2008), p. 35.
[36] Strawson (2002), pp. 457-8.
[37] Denying both premise (1) and (3) as libertarians do requires affirming the truth of indeterminism, which is precisely why we are able to rule these possibilities out. Clarke (2005: p. 16), for instance, denies the claim in premise (1) that how you act is a function of the way you are, saying: "The agent might act freely […] because it might be up to him whether his being a certain way is followed by his acting as he does, and so be up to him whether if he is that way mentally, then he so acts." Kane,

this section will focus on the compatibilist's objection to premise (2), which finds its fullest articulation in a paper by Hurley in which she asks: 'Is Responsibility Essentially Impossible?'[38]

Hurley sums up her objection to Strawson's argument as follows: "According to one natural set of intuitions, a person need not be responsible for being what he is in order to be responsible for choices that are determined by what he is."[39] On this understating of the conditions for moral responsibility, the buck must stop somewhere, and as a matter of fact it stops at our characters. That is to say, even though it is true that we are not responsible for our characters in the way Strawson demands we must be in order to be responsible for our actions, it is simply wrong to insist that responsibility for one's character is a condition for responsibility for one's actions. To put the point in a way that relates it to Strawson's argument, Hurley thinks we might deny the truth of the conditional expressed in premise (2), and thus reject the argument as a whole. It is not the case (runs the objection) that we cannot be ultimately morally responsible for what we do despite the fact that we are not—and indeed never could be—ultimately morally responsible for the way that we are.

Is Hurley right to think that it is natural to intuit that we can be morally responsible for our actions, even given the belief or knowledge that we are not responsible for our character? Nozick believes so, offering his opinion to the effect that we can be deserving of praise and blame for our actions without necessarily being responsible for the character traits and other factors that explain those same actions:

> It is not true, for example, that a person earns *Y* (a right to keep a painting he's made, praise for writing *A Theory of Justice*, and so on) only if he's earned (or otherwise *deserves*) whatever he used (including natural assets) in the process of earning *Y*. Some of the things he uses he just may *have*, not

---

meanwhile, denies the assertion in premise (3) that you cannot be ultimately responsible for the way you are by citing the Self Forming Actions (SFAs) that we discussed in the previous chapter: these allow us to wrest control from determining forces and make us the originators of our own character and actions.

[38] Others have articulated much the same objection. For example, Mele (1995: p. 224) worries: "It is being claimed, in effect, that the very definition of 'true responsibility' entails that possessing such responsibility for any choice requires having made an infinitely regressive series of choices." While I shall focus discussion on Hurley's criticisms of Strawson, she and Mele share a similar basic worry with Strawson's account of moral responsibility, which is that he defines the concept of moral responsibility in such a way that its instantiation requires the impossible to occur (namely, for an infinitely regressive series of choices to be made).

[39] Hurley (2000), p.30.

illegitimately. It needn't be that the foundations underlying desert are themselves deserved, *all the way down*.[40]

This at least shows that Hurley is not alone in having the intuition that premise (2) might not be true, and that we might in fact be responsible for our actions without always being responsible for all of the elements of our character from which these actions find their source. Still, all that has been presented so far, in the form of Hurley's and Nozick's assertions, is evidence that people sometimes intuitively believe the falsity of premise (2). More will be required if we are to demonstrate that Hurley's and Nozick's intuitions about the falsity of premise (2) are actually correct.

To this end, we should consider Hurley's rejection of Strawson's premise (2) in the context of the project of the paper in which it appears. As has been mentioned, Hurley's paper attempts to answer the question of whether true moral responsibility is essentially impossible, a question which Strawson of course answers in the affirmative. Early on in her paper, Hurley muses: "suppose a theory requires for responsible action that people control not just their actions but the causes of their actions, and their causes in turn, all the way back."[41] This is just the sort of requirement for responsible action which premise (2) of Strawson's argument demands, and which it appears cannot be met. And now, since this condition can be seen to be non-empirically impossible to satisfy, the following dilemma arises: "Does this show that all [ascriptions of responsibility have been made] in error, that no one is ever responsible for anything? Or does it show that the theory is in error, that it misdescribes the conditions for responsible acts?"[42]

As may be guessed, Hurley plumps for the latter solution to the dilemma, arguing that so-called 'error theories' regarding moral responsibility are themselves (rather ironically, it may be thought) in error. The dilemma posed is one between adopting a position of eliminativism regarding moral responsibility (as Strawson does), or else favouring a revisionist position (as Hurley does)—revisionist in the sense that, if we had previously assumed that moral responsibility must involve a commitment to

[40] Nozick (2012), p. 225.
[41] Hurley (2000), p. 3.
[42] *Ibid.*, p. 3.

what Hurley terms a 'regression requirement' (a requirement to which premise (2) is committed), then this assumption will need to be revised.[4344]

Hurley argues that there are good reasons to favour the revisionist stance over an eliminativist one. Perhaps the main reason she cites is that eliminativists such as Strawson are required to pull off what she inventively terms an "interpretative double whammy": they must persuade us that it is both analytically impossible for any agent to fulfil the regression requirement demanded by premise (2), and also necessary that they do so in order to be ultimately morally responsible.[45] So, while some varieties of eliminativism make the weaker claim that some property P that is essential to kind F is as a matter of fact uninstantiated and thus that there are no Fs, others make the much stronger claim that it is *impossible* that property P, which is essential to kind F, should be instantiated. Galen Strawson's position, according to which "true self-determination is both necessary and logically impossible," is an example of the latter variety of eliminativism, which Hurley duly dubs 'impossible-essence eliminativism.'[46]

This rejection of the regression requirement in premise (2) is, in Hurley's view, a natural corollary of her favouring what she terms a "context-driven approach" to the issue of moral responsibility. Such an approach stands in contrast to theory-driven approaches to moral responsibility in the following respect: while theory-driven approaches involve satisfying certain conditions which, collectively, can be regarded as defining a theoretical role for moral responsibility, context-driven approaches instead focus on how the notion of moral responsibility is applied. Strawson's demand that moral responsibility satisfies the regression requirement condition thus marks out his approach as being theoretically driven, while Hurley emphasises the importance of focusing on contexts of use. In Hurley's words, a context-driven

[43] Hurley (2000: p. 24) defines this 'regression requirement' by saying: "to be responsible for something you must be responsible for its causes, and it applies recursively." Although worded slightly differently from premise (2), it seems to me that both express the same intuition, and therefore both either stand or fall together.
[44] Of course, that Hurley labels her position 'revisionist' does not imply that she thinks all will consider it a revision. In fact, and as just noted, Hurley makes the case that the intuition that premise (2) is false is reasonably commonplace, citing Nozick as one example of what we might call (with no disparagement intended) "regression requirement deniers." The point is simply that, insofar as we have a tendency to either tacitly assume or else explicitly acknowledge the truth of the regression requirement, Hurley's position constitutes a revision to our notions of moral responsibility.
[45] Hurley (2001), p. 21.
[46] Strawson (2010), p. 50.

account "gives explanatory priority to contexts of positive use, though it can admit that some of these are mistaken." [47]

We can now see why Hurley is unprepared to accede to the truth of Strawson's eliminativism—because, as she says: "the revisionist who invokes a context-driven account cannot allow that none of our positive applications of 'responsibility' hit their mark, since some such must anchor her claim to be talking about responsibility."[48] Hurley is claiming that if one is inclined (as she is) to adopt a context-driven account, then it cannot be the case that none of one's attributions of responsibility are correct. If all attributions of responsibility turned out to be false, Hurley argues, then proponents of context-driven accounts would not even be talking about responsibility. And since they patently are talking about responsibility, Hurley rules that it is extremely unlikely that eliminativism regarding moral responsibility is correct.

4.5.1 Response to Hurley

Hurley has argued that the notion of an impossible essence is at best suspect and at worst meaningless. Her context-driven approach to the issue of moral responsibility favours revisionism as opposed to eliminativism and, while she stops short of definitively ruling against the latter, her conclusion is that Strawson's argument is uncharitable in the extreme in choosing to interpret ascriptions of moral responsibility as involving incoherence. Strawson is unperturbed by this line of criticism, however, and he defends his commitment to the claim that our attributions of moral responsibility are incoherent. This he does by pointing out that the putative incoherence of moral responsibility attributions does not entail a commitment to the belief that the concept of moral responsibility is meaningless. On the contrary, Strawson wishes to affirm that moral responsibility is meaningful even while he denies that it is even possible for any agent ever to be truly morally responsible. He writes:

> [I]t is precisely (only) because one has a grasp of the content of the notion of [ultimate moral responsibility] that one can see, or can be brought to see, that it is incoherent. It is the same with the

[47] Hurley (2001), p. 39.
[48] *Ibid.*, p. 38.

notion of a round square. Some may say that they don't really know what the content of this notion is, but it is easy to specify. A round square is an equiangular, equilateral, rectilinear, quadrilateral closed plane figure every point on the periphery of which is equidistant from a single point within its periphery. It is because we know the content of the notion that we know that there cannot be such a thing as a round square, and the same is true of the notion of [ultimate moral responsibility]. Many say that statements or concepts that are self-contradictory are meaningless, but meaningfulness is a necessary condition of contradictoriness.[49]

I think that Strawson is right to defend the claim that his arguing that all attributions of moral responsibility are incoherent does not entail that the very notion of moral responsibility is meaningless. That moral responsibility meets the regression requirement is part of the essence of moral responsibility, Strawson wishes to claim, and the fact that this requirement can never be met does not render the concept of moral responsibility meaningless. Even while remaining agnostic regarding the truth of whether there is a regression requirement on moral responsibility (at least for the time being), I think it is clear that affirming such a requirement does not render the concept of moral responsibility meaningless, just as the notion of a round square is both impossible to instantiate and not meaningless.

Another issue to examine is context-driven accounts of moral responsibility. In particular, we can ask whether Hurley is right to adopt such an account, and also whether she is right to think that its adoption effectively blocks any claim of eliminativism regarding moral responsibility. Dealing firstly with the question of Hurley's rationale for favouring a context-over a theory-driven approach in the case of moral responsibility ascriptions, we can note how she observes that the latter are "attractive where revisionism seems intuitively correct."[50] For example, she explains, it might have been part of our ancestors' beliefs that stars are the spirits of the great kings of the past. As such, if we were to adopt a theory-driven approach to stars whereby it is an essential property of these objects that they are the spirits of the great kings of the past, we would thereby be committed (implausibly, of course) to eliminativism about stars. Context-driven accounts, by contrast, which protect applications of a term at the expense of extant theory, would allow for certain misattributions of the term 'star' (since, for example, we might erroneously apply the term to planets until we realise that planets lack certain essential properties of stars)

[49] Strawson (2002), p. 452.
[50] Hurley (2000), p. 12.

but nevertheless ensure that some of our attributions are correct. Just as a context-driven approach is attractive vis-à-vis stars, so it is for moral responsibility: it favours revisionism over eliminativism, which seems intuitively correct in both cases.

I am not convinced of the reasonableness of Hurley's rationale for endorsing a context-driven approach. It sounds rather as though Hurley has made up her mind on the issue of eliminativism regarding moral responsibility before even considering Strawson's argument, and that she therefore allows prior intuitions about the appropriateness of ascribing moral responsibility to guide her thinking rather than engaging with the logic of the argument. Tellingly, nowhere in Hurley's paper does she offer any explanation as to exactly how Strawson's regression requirement is misguided (that is, aside from an appeal to her and Nozick's intuitions on the issue, she offers no argument for rejecting the claim that, to be morally responsible for what you do, you must be morally responsible for the way you are).

In any case, it is not at all clear that adopting a context-driven approach blocks the move to eliminativism: I see no reason why one should not allow for the possibility that *all* of what Hurley terms "contexts of positive use" are mistaken.[51] That is, adopting an approach towards analysing moral responsibility whereby, as Hurley recommends, you begin by examining contexts of use for the term and then build a theory around these positive attributions, does not seem to rule out the possibility of subsequently discovering that all positive applications of the term are wrong. Certainly, having an intuition in favour of retaining the concept of moral responsibility is sure to influence (or, less neutrally, bias) your judgements regarding theory; but having these intuitions and adopting a context-driven account as a result will not to any degree help to establish the falsity of eliminativism.

After considering the arguments in detail, I conclude that Strawson is right to insist on the regression requirement expressed in premise (2) of his argument. While Hurley contends that one natural set of intuitions leads us to believe that a person "need not be responsible for being what he is in order to be responsible for choices that are determined by what he is," a much more natural intuition to my mind is to believe the exact opposite: that is, that a person *cannot* be responsible for choices

[51] *Ibid.*, p. 39.

that are themselves determined by something for which he is not responsible.[52] Further, Hurley's contention receives no support from her attempts to establish that the notion of an impossible essence—which Strawson must characterise moral responsibility as being—is in some way suspect or even meaningless. It is neither of those things: on the contrary, there can be no difficulty in understanding Strawson's insistence that the regression requirement is both a necessary condition for moral responsibility and one that is impossible to satisfy.

4.6 Conclusion

This chapter has expounded and critiqued what I consider to be the two most persuasive arguments for hard determinism. I have argued that even the best and most ingenious compatibilist attempts to block these arguments' conclusions are unsuccessful, and that there is therefore a strong case for accepting the hard determinist's contention: no-one has the free will required for moral responsibility. Nevertheless, despite my considering them persuasive, it must be acknowledged that they fail to win over the numerous philosophers who maintain that determinism and the free will required for moral responsibility are compatible. The question then is: is there any one issue in particular that lies at the root of this disagreement? I would argue that there is, and it is an issue, moreover, that we encountered previously when discussing libertarianism: the origination condition. The notion of origination was invoked not only in Galen Strawson's explanation of what it is to be ultimately responsible but also in Kane's elucidation of his libertarian theory of free will, who nicely encapsulates it as follows:

> This, of course, is the governing image of free will that we tried to spell out in terms of UR: the origins or sources, or *archai*, of actions should be in the agents themselves and not in something outside the agents, if the agents are to be ultimately responsible for what they do and what they are.[53]

It can be seen how the origination condition plays a role in both van Inwagen's and Galen Strawson's respective arguments, since both have at their core the idea that an agent's supposedly freely-willed actions are a consequence of something over which the agent has no control. Not being in the agent's control, the agent is therefore not

[52] *Ibid.*, p.30.
[53] Kane (1996), p. 70.

responsible for the action. In van Inwagen's case, that 'something' over which the agent has no control is the fact that the state of the world at some prior time together with the laws of nature entail one's own actions (as expressed in premise (2) of his argument), whereas for Galen Strawson, it is the character from which an agent's actions flow that he cannot control and hence for which he cannot be held responsible (also expressed in premise (2) of his argument).

So we can see that the condition of origination concerns the sources of our actions, and is the requirement that these should come from the agent rather than from some external source over which the agent lacks control. While van Inwagen's and Galen Strawson's arguments both rely on the intuition that the origination condition ought to be met, the compatibilist, on the other hand, must deny this. I do not wish to suggest that the case against those compatibilists who seek to deny the conclusions of van Inwagen's and Galen Strawson's hard determinist arguments is conclusive. However, I do say that the examples considered in this chapter lend strong support to the view that the origination condition is indeed a necessary condition for true moral responsibility.

5

Mixed Views on Moral Responsibility

Having rejected libertarianism, and having judged that there are strong arguments for rejecting compatibilism also, we continue with our task of seeking to identify the correct theory of moral responsibility. Libertarianism was rejected on the grounds that the truth of the *PSR* entails determinism, which in turn entails the falsity of libertarianism. Compatibilism, meanwhile, was strongly called into question on account of the arguments of van Inwagen and Galen Strawson. The common thread between these arguments is that they are concerned with one's actions being beyond one's control, of not originating with the agent. Compatibilists argue in various ways that this lack of origination does not impact upon one's free will or moral responsibility. The responses to van Inwagen's and Strawson's arguments, however, are not able to establish this convincingly.

Harking back to the beginning of the previous chapter, a table was presented to illustrate the various permutations of opinion that one can have regarding determinism and freedom. At this stage, we find ourselves inclined towards position (1) from that table, which is the hard determinist's position: determinism is true, and this is not compatible with the free will required for moral responsibility. In fact, if we are convinced that we can discount libertarianism and compatibilism, then we have no choice but to adopt the hard determinist position.

However, there are many who view hard determinism as such an unappealing position to take that it would be better avoided at all costs.[1] It is considered an

[1] For example, van Inwagen (1983: p. 207) argues that the reality of moral responsibility is too self-evident to deny, and he illustrates his argument with the following humorous example: "I have listened to philosophers who deny the existence of moral responsibility. I cannot take them seriously. I know a philosopher who has written a paper in which he denies the reality of moral responsibility. And yet this same philosopher, when certain of his books were stolen, said, "That was a *shoddy* thing

unappealing position for various reasons. For one thing, it is sometimes argued that the belief in one another's moral responsibility, at least some of the time, is an inescapable part of our thinking.[2] This belief of course translates into the practices that we have in our society, such as blaming those who commit offences and sanctioning them by various means, as well as praising and honouring those who are considered deserving by virtue of the good deeds that they have performed. Embracing hard determinism, therefore, would entail transforming our beliefs concerning moral responsibility, and in all likelihood would also entail the alteration of our attitudes and practices. Undergoing such a transformation in our beliefs, attitudes and practices would be intolerable, impracticable, and perhaps even impossible (or so the argument goes).

And so it is in the spirit of attempting to avoid this hard determinist position that we turn now to what can be called 'mixed views' on moral responsibility! Mixed views derive their name from the fact that they combine elements of compatibilist and incompatibilist thought (with libertarianism and hard determinism being the two species of incompatibilism). Their appeal lies in the fact that they hold out the promise of providing a more satisfying response to the problem of origination than we encountered from standard compatibilist responses to van Inwagen's and Galen Strawson's arguments, while allowing us to avoid the all-out denial of the reality of moral responsibility that embracing hard determinism seems to demand. Two mixed views will be under the spotlight in this chapter, the first of which is expounded by Manuel Vargas and goes by the name of 'Revisionism,' while the second is Richard Double's theory of 'Metaethical Subjectivism.' An account of each of these theories will be given, with the hope being that insights from either or both of them might help us when it comes to constructing a positive account of moral responsibility—an account that we hope can avoid the pitfalls of standard compatibilist and hard determinist theories.

---

to do! But no one can consistently say that a certain act was a shoddy thing to do *and* say that its agent was not morally responsible when he performed it: those who are morally responsible for what they do may perhaps deserve our pity; they certainly do not deserve our censure."

[2] P.F. Strawson's (1963: p. 9) highly influential article "Freedom and Resentment" provides the classic statement of the belief that a theoretical conviction that determinism is true poses no threat to our current moral responsibility practices. P.F. Strawson writes: "[W]e cannot, as we are, seriously envisage ourselves adopting a thoroughgoing objectivity of attitude to others as a result of theoretical conviction of the truth of determinism."

5.1 Vargas's Revisionism

Revisionism with respect to theories and practices surrounding moral responsibility comes in many forms. What all of these forms have in common, however, is that they involve noting a discrepancy between two different projects, one of which can be called *diagnostic*, and the other *prescriptive*. So, all revisionist theories will on the one hand offer a diagnosis of our common sense opinions regarding moral responsibility. This would explain what we (in the main) *in fact* believe to be true regarding responsibility, and most crucially the conditions under which an agent is thought to be morally responsible. On the other hand, revisionist theories seek also to prescribe, to make the case for what we *should* believe to be true regarding responsibility.[3] It is this positing of a discrepancy between what we in fact believe and what we ought to believe, that makes any given theory a revisionist theory.

Vargas identifies himself as a 'moderate revisionist.' A wide range of views come under the umbrella of moderate revisionism, Vargas informs us, which is perhaps not surprising given his broad definition of it, which is as follows: "moderate revisionism is the idea that the folk concept of responsibility is inadequate until it has been modified in some way."[4] Still, such a definition does distinguish moderate revisionism from both weak and strong varieties in that it reveals that moderate revisionism is not limited to mere clarifications of linguistic or conceptual confusions (in contrast to weak revisionism), yet neither does it advocate the straightforward elimination of the concept of responsibility, nor of related attitudes and practices (in contrast to strong revisionism). Instead, what it calls for is a "pruning," to use Vargas's term, of one or more of these elements. While this might involve the elimination of aspects of these elements, there is no wholesale elimination of the elements themselves.

[3] Note that, according to Vargas's (in Fischer *et al.* (2007), p. 151) account of revisionist theories, the sole criterion for judging whether a theory is revisionist is whether it prescribes different *beliefs* from those it diagnoses. If it does, then it is revisionist; if not, then it is not. As for attitudes and practices, while Vargas sees these (along with beliefs) as elements of our thought and behaviour that require justification on account of the fact that they reveal our attributions of free will and moral responsibility, nevertheless, and strictly speaking, the concept of revisionism is defined solely in relation to our beliefs—that is, a theory is revisionist if and only if it claims that our beliefs are different from what they should be.

[4] Vargas (2005), p. 409.

How does Vargas propose that we modify our folk concept of responsibility? He suggests we do so by pruning away the libertarian beliefs that he detects in us:

> [W]e see ourselves as having genuine, robust alternative possibilities available to us at various moments of decision. We may even see ourselves as agent-causes, a special kind of cause distinct from the non-agential parts of the causal order. Moreover, we tend to think of this picture of our own agency as underwriting many important aspects of human life, including moral responsibility.[5]

Vargas's diagnosis is that our folk beliefs are libertarian: most people have a propensity towards a pre-philosophical belief in robust alternative possibilities, of a sort whose veracity would entail the truth of indeterminism in the realm of human action. It might further be the case, Vargas adds, that the majority of us are predisposed to believe that we are agent-causes, set apart from the causal order. These folk metaphysical beliefs provide the grounding for our belief in moral responsibility: that is, we believe that we are morally responsible *because of* our metaphysical intuitions.[6] This conception of ourselves and of what is required for us to be morally responsible is, continues Vargas, "implausible and largely unnecessary."[7] To the extent that we hold a libertarian picture of human agency, our folk concept of moral responsibility is in error. Despite his insistence that our folk concept is in error, when it comes to prescription—that is, his account of what we *should* believe regarding moral responsibility—Vargas argues that we need not jettison talk of moral responsibility, praise, and blame. On the contrary, these properties really are instantiated, and so the prescriptive part of his revisionism assures us that we should expect success in locating them.

What we have when we put together the descriptive and prescriptive parts of Vargas's theory is a hybrid account, as Vargas himself acknowledges: it is one that is incompatibilist in its diagnosis (of what we *in fact* believe), but compatibilist in its prescription (for what we *should* believe). We are inclined to see things, one might

---

[5] Fischer *et al.* (2007), p. 128.

[6] Vargas is far from alone in believing that most people are predisposed to favouring a libertarian position on the issue of free will. For example, Pereboom (2001: xvi) states: "Beginning students typically recoil at the compatibilist response to the problem of moral responsibility"; Ekstrom (2002: p. 310) notes: "[W]e come to the table, nearly all of us, as pretheoretic incompatibilists"; and, lastly Kane (1999: p. 219) declares: "In my experience, most ordinary persons start out as natural incompatibilists […] Ordinary persons have to be talked out of this natural incompatibilism by the clever arguments of philosophers."

[7] Fischer *et al.* (2007), p. 128. Vargas (2011: p. 20) also declares: "If the integrity of our normative practices rested on [libertarianism], it would leave us in the morally precarious situation of blaming and punishing people on the basis of a fervent hope that we have it."

say, from a libertarian perspective;[8] but we are enjoined to recognise that we must be compatibilist in our outlook instead. When it comes to this compatibilist prescription, a satisfactory moderate revisionism of the type that Vargas envisages will need to provide accounts of two distinct concepts: (1) of responsible agency, and; (2) of responsibility norms.

Vargas summarises his own account of (1) as follows: "On the particular account I favour, the distinctive mark of the freedom-relevant aspects of responsible agency is the agent's sensitivity to specifically moral considerations and the capacity of that agent to appropriately govern his or her conduct in light of those considerations."[9] Vargas elaborates on this by explaining that considerations are the kinds of things that can generate reasons; thus, to be a responsible agent is to be able specifically to detect *moral* considerations, and so to have reasons for acting that are generated by these moral considerations.

This account of the concept of responsible agency is not a novel one. As Vargas acknowledges, Fischer and Ravizza put forward a very similar position with their reasons-responsive account. What does make Vargas's position distinctive, however, is his account of (2), the concept of responsibility norms. More importantly, at least for our purposes, is the fact that Vargas's account is distinctive in a way that might prove useful in the current dialectical situation: that is, it offers the hope of providing justification for our "responsibility norms"—essentially, our practices of praising and blaming—without needing to deny the insights encapsulated in the arguments of van Inwagen and Galen Strawson. Vargas makes possible this compatibilism (for want of a better word!) between the arguments for hard determinism, and the insistence that we can and should maintain our current responsibility norms, by arguing that these norms are the way they are for utilitarian purposes. The following passage provides a sketch of how Vargas sees this working:

> When you judge me blameworthy for being insensitive to someone's feelings, the sting of your disapproval forces me to attend to considerations that I might have failed to see or failed to act on in the right way. Over time, and given widespread participation in this system of judgments, practices,

[8] Vargas (in Fischer *et al.* (2007: p. 152)) states that the libertarian perspective from which we are inclined to see things includes "minimally, metaphysically robust alternative possibilities."
[9] Fischer *et al.* (2007), p. 155.

and attitudes we come to help both ourselves and other consideration-sensitive creatures to better track what moral considerations there are.[10]

Our responsibility norms thus serve an essentially utilitarian function, that function being to improve us morally over time. That is not to say that that is *all* they are; neither is it to say that this is how we generally perceive them either—nonetheless, Vargas contends that responsibility norms are essentially utilitarian in function.[11] Additionally, Vargas insists that it is precisely because responsibility norms serve this utilitarian function that we are justified in acting in accordance with them. It seems possible to read Vargas as declaring that utility is in fact the sole justification for responsibility norms, since he claims that these norms are only justified "inasmuch as they, on the whole and over time, tend to contribute to our better perceiving and appropriately responding to moral considerations."[12] However, in another passage, Vargas appears to adopt a less strident position in allowing that, while we can justify the bulk of our responsibility-characteristic practices and attitudes with reference to their utility-maximising qualities, this "does not preclude other ways of justifying the responsibility system or parts of it."[13] Vargas does not elaborate on this or offer further speculation as to how else we might justify our responsibility norms besides suggesting that there may be overlapping justifications available. It may turn out to be the case that utility is not the sole justification for our responsibility norms, then, although Vargas insists that it is the primary justification.

Before moving on to a full critique of Vargas's moderate revisionism, a final consideration to note is that he does not believe that his approach binds him to accepting all of our current responsibility norms. On the contrary, Vargas counsels that it would be unduly optimistic to imagine that our current norms just so happen to be the "normatively ideal norms."[14] It is more plausible to think instead that most of our responsibility-characteristic practices, attitudes, and beliefs can be justified on utilitarian grounds, although there is no guarantee that even this much is so.

---

[10] *Ibid.*, p. 155.
[11] To apply Vargas's terminology to his own theory, it seems fair to characterise his position on our perception of responsibility norms as revisionist. That is, Vargas thinks that, although we do not generally perceive responsibility norms as being essentially utilitarian in function, this is what they in fact are.
[12] Fischer *et al.* (2007), pp. 155-6.
[13] *Ibid.*, p. 156.
[14] *Ibid.*, p. 156.

5.2 Debating Vargas's Diagnosis

As has been explained, Vargas's moderate revisionism has two facets: the diagnostic and the prescriptive. Concerning the diagnostic facet, Vargas identifies a strain of libertarianism in our folk intuitions surrounding moral responsibility, which leads him to believe that libertarian freedom seems worthwhile "at least because it is the only kind of theory that preserves our ordinary concept of responsibility."[15] Elsewhere, however, Vargas has acknowledged that this diagnosis could be incorrect, such as in the following passage:

> Most current revisionist accounts have been revisionist compatibilisms, motivated by the conviction that folk beliefs contain incompatibilist elements. It would be problematic if it turned out that ordinary persons did not have incompatibilist commitments.[16]

A first criticism is that Vargas is quite right to raise the possibility that ordinary persons do not have incompatibilist commitments of the sort that he believes them to have: rather, most of us are natural born compatibilists, so this criticism goes, and therefore Vargas's prescription of compatibilism is no revision at all. Evidence in favour of the claim that we are in fact folk compatiblists rather than folk libertarians comes in the form of a paper by Nahmias *et al.*[17] In order to challenge the claim that incompatibilism is intuitive, Nahmias *et al.* tested the following prediction, which folk incompatibilists, Vargas included, should expect to be found accurate:

(P) When presented with a deterministic scenario, most people will judge that agents in such a scenario do not act of their own free will and are not morally responsible for their actions.[18]

The prediction (P) was put to the test by surveying people who had not studied the free will debate. These people were presented with the following scenario, which draws on a Laplacean conception of determinism:

> Imagine that in the next century we discover all the laws of nature, and we build a supercomputer which can deduce from these laws of nature and from the current state of everything in the world

[15] Vargas (2004), p. 237.
[16] Vargas (2010), p. 25.
[17] Nahmias *et al.* (2006). For studies that likewise find that the majority of people are natural compatibilists, see also Viney, Waldman and Barchilon (1982); and Woolfolk, Doris and Darley (2006).
[18] Nahmias *et al.* (2006), p. 36.

exactly what will be happening in the world at any future time. It can look at everything about the way the world is and predict everything about how it will be with 100% accuracy. Suppose that such a supercomputer existed, and it looks at the state of the universe at a certain time on March 25th, 2150 A.D., twenty years before Jeremy Hall is born. The computer then deduces from this information and the laws of nature that Jeremy will definitely rob Fidelity Bank at 6:00 PM on January 26th, 2195. As always, the supercomputer's prediction is correct; Jeremy robs Fidelity Bank at 6:00 PM on January 26th, 2195.[19]

Once participants had been presented with this scenario, Nahmias *et al.* asked: "Do you think that, when Jeremy robs the bank, he acts of his own free will?" A significant majority (76%) of participants judged that Jeremy does act of his own free will.[20] In order to assuage worries that people might be more inclined to overlook mitigating factors when an agent performs an act which is deemed to be immoral, Nahmias *et al.* asked another set of participants to judge a similarly Laplacean scenario, but one in which the action performed was positive rather than negative. Instead of robbing a bank, the scenario on this occasion involved Jeremy saving a child. A third scenario, meanwhile, presented to a third set of participants, involved the neutral act of Jeremy going jogging. The results showed that changing the nature of the action had no significant effect on responses: 68% and 79% of participants respectively judged that Jeremy acted of his own free will. Additional sets of participants were also asked directly for their opinion on Jeremy's moral responsibility: 83% judged that Jeremy is "morally blameworthy for robbing the bank," and 88% judged that "he is morally praiseworthy for saving the child."[21]

Similar deterministic scenarios were also presented to participants. For instance, they were asked to imagine one scenario in which the universe was recreated over and over again, starting from the exact same initial conditions and with the same laws of nature, such that each time the same woman, Jill, decides to steal a necklace at the same time. Another scenario made salient the fact that the agents' actions (twin

[19] Nahmias et al (2006), p. 36. Note that the word 'determinism' is not used in any part of the description of this scenario. Nahmias *et al.* (2006: p. 37) write that the reason for this is that prior surveys showed that "most people either did not know what 'determinism' meant or they thought it meant, basically, the opposite of free will." Using the term 'determinism' would therefore be of no use for those who did not know the meaning of the term, while it would be positively prejudicial in the case of those who thought it meant the opposite of free will, since incompatibilism would be judged by these people to be correct by definition, regardless of what intuitions they might have when faced with actual deterministic scenarios.

[20] There is a complete description of this and similar studies, including the methodology used, in Nahmias *et al.* (2005).

[21] Nahmias *et al*. (2006), p. 37.

brothers, in this case) were deterministically caused by factors outside their control, namely their genes and upbringing: one twin, Fred, having been adopted by the selfish Jackson family has been caused to value money above all else and to think that it is acceptable to acquire money in any way you can; the other twin, Barney, has been caused through a combination of genetic inheritance and upbringing in the kindly Kinderson family, to value honesty above all else and to respect the property of others. When each man one day finds a wallet containing $1000 and an ID card, only Barney returns the wallet to the owner, while Fred (somewhat predictably) trousers the money.

Participants were asked the same questions of the agents in each of the above scenarios, namely: did they act of their own free will, and; were they morally responsible for their action? In all cases, the majority of participants judged that the agent or agents acted of their own free will and were morally responsible, thereby offering strong evidence for the falsity of (P), the incompatibilist prediction that most ordinary people would judge that agents in a deterministic scenario lack free will and moral responsibility for their actions.[22]

The implications for Vargas's proposal that we should adopt his form of moderate revisionism should be clear: if the findings of Nahmias *et al.* are correct, then it is not the case that the majority of ordinary people have an incompatibilist folk theory of responsibility. These findings put the diagnostic element of Vargas's theory in jeopardy, and compel him to respond in one of two ways: he must either argue that there is some alternative way of understanding the claim that the majority of us are folk incompatibilists, one that, crucially, does not commit him to accepting (P); or else he must show that there is some flaw in the methodology employed by Nahmias *et al.*, a flaw that precludes them from concluding the falsity of (P).

[22] A table displaying the summary of results from all three scenarios is presented in Nahmias *et al.* (2006: p. 39) and reproduced below:

Subjects' judgments that the agents…	Scenario 1 (Jeremy)	Scenario 2 (Jill)	Scenario 3 (Fred & Barney)
…acted of their own *free will*	76% (robbing bank) 68% (saving child) 79% (going jogging)	66%	76% (stealing) 76% (returning)
…are *morally responsible* for their action	83% (robbing bank) 88% (saving child)	77%	60% (stealing) 64% (returning)

In truth, pressure can be applied to both horns of this dilemma. Let us begin by looking at how Vargas can tackle the second horn, which requires him to find some flaw in Nahmias *et al.*'s methodology, thereby undermining their claim that (P) is false.[23] Sarkissian *et al.* (2010) highlight one methodological flaw in particular. They point out that the kind of concrete, affect-laden cases that Nahmias *et al.* concoct are just the kinds of cases that may (according to a wealth of studies in social psychology) introduce biases in folk judgments. It remains to be seen, therefore, whether "compatibilist intuitions hold up when participants are presented not with a case likely to trigger affect, but instead asked more directly whether moral responsibility can be possible [*sic*] in a deterministic universe."[24]

Nichols and Knobe (2007) sought to resolve just this issue in designing the following experiment. Two Universes are described to participants, Universe A and Universe B. Whereas in Universe A everything is caused by preceding events (i.e. it is a determined universe) in Universe B everything *apart from human decisions* is caused by preceding events. Once participants have grasped this difference between the two universes, they are randomly assigned to one of two groups: the concrete condition or the abstract condition. Those assigned to the concrete condition are given the following question:

In Universe A, a man named Bill has become attracted to his secretary, and he decides that the only way to be with her is to kill his wife and 3 children. He knows that it is impossible to escape from his house in the event of a fire. Before he leaves on a business trip, he sets up a device in his basement that burns down the house and kills his family.

Is Bill fully morally responsible for killing his wife and children?

YES NO

No scenario is presented to those in the abstract condition, who are instead simply asked the question:

In Universe A, is it possible for a person to be fully morally responsible for their actions?

YES NO[25]

[23] For critiques of Nahmias *et al.*, see Sarkissian *et al.* (2010); Feltz, Cokely and Nadelhoffer (2009); Roskies and Nichols (2008); Misenheimer (2008); and Nichols and Knobe (2007).
[24] Sarkissian *et al.* (2010), p. 347.
[25] Nichols and Knobe (2007), p. 670.

Nichols and Knobe found that, for the concrete condition, 72% of participants responded that Bill was fully morally responsible, a result that appears to support the view that the majority of people's intuitions regarding moral responsibility are compatibilist. However, when it came to the abstract condition, 86% of participants responded that it is not possible to be fully morally responsible in Universe A, lending support to an incompatibilist stance on folk intuitions. These findings give credence to the hypothesis that affect plays a key role in generating people's compatibilist intuitions. The thought now is that these results might provide some ammunition for Vargas against Nahmias *et al*., as they offer evidence that the majority of us are indeed folk incompatibilists - at least when presented with a scenario which is not affect-laden.

Related to this issue of affect-laden, concrete examples versus more neutral, abstract ones, is the issue of where the presented scenario is set. Roskies and Nichols (2008) noted that many of the scenarios presented by Nahmias *et al.* were supposed to occur in our own world, whereas the scenarios presented by Nichols and Knobe were always set in an alternate universe. Conducting their own experiment, Roskies and Nichols randomly assigned participants to two conditions: *Actual* and *Alternate*. In both conditions, subjects were provided with a sketch of a deterministic world, the only difference being that, in the *Actual* condition, this universe was clearly implied to be our own, whereas in the *Alternate* condition ("Universe A") the universe was explicitly not ours. In each case, eminent scientists were said to have discovered beyond any reasonable doubt that all events, including human actions, are determined. Participants were then presented with the following statement to match their condition, and asked to rate their level of agreement (from 1 [disagree completely] to 7 [agree completely]):

- *Alternate*: If these scientists are right, then it is impossible for a person in Universe A to be fully morally responsible for their actions.
- *Actual*: If these scientists are right, then it is impossible for a person to be fully morally responsible for their actions.

The results were striking: the mean response in the *Alternate* condition was 5.06, against a mean response in the *Actual* condition of 3.58. So, participants were much more inclined to give compatibilist responses when they were asked to assume that

determinism was true of their own world, as opposed to when determinism was operating in some other world.[26]

Do the findings of Nahmias *et al.*'s critics suggest that prediction (P) is tenable after all? I would argue they do not. What they undoubtedly do show is that intuitions regarding free will and moral responsibility vary according to the details of the deterministic scenario sketched; but (P) predicts that most people have the folk intuition that determinism precludes the possibility of free will and moral responsibility, when the fact is that that blanket assertion is not true. Further, it should be remembered that Vargas is in any case concerned to capture our folk intuitions regarding *this* world, and thus we should expect him to be more interested in the findings from Nahmias's *et al.*'s concrete, affect-laden examples than the more abstract, affectless ones of Nichols and Knobe, or Roskies and Nichols. There is no dodging the conclusion, then, that (P) is false.

While the first horn of the dilemma can be ruled out, the second horn remains to be considered. This second horn sees Vargas conceding that (P) is false, while at the same denying that the diagnostic element of his moderate revisionism commits him to believing that (P). Endorsing the following alternative prediction to (P) could enable Vargas to achieve that aim:

(P*) When presented with scenarios in which human actions are either deterministic or indeterministic, most people will judge: (a) that our universe is most like the indeterministic scenario, and; (b) that it is possible for a person to be fully morally responsible in our universe.

Prediction (P*), as can be seen, is in fact an amalgam of two separate predictions.[27] The first of these is that most people will judge that human actions are not determined in our universe, while the second makes the prediction that most people will judge that full moral responsibility is a feature of our universe. If it turns out that

[26] Roskies and Nichols (2008), pp. 373-4.

[27] A third prediction could usefully be added, which is that people judge that indeterminism is compatible with full moral responsibility. This would rule out the possibility that people might endorse both (a) and (b), and yet not be conscious that these two judgments conflict with a further belief in the incompatibility of indeterminism and full moral responsibility. Unfortunately, this prediction remains untested, so I note this just to flag up the possibility of people holding conflicting beliefs. I am happy to proceed, however, on the assumption that the great majority of people hold no such conflicting beliefs.

not only are these predictions correct, but that the vast majority of people endorse both (a) and (b), then it follows that most people think that the libertarian picture of free will is the correct one, at least when it comes to our own world. In endorsing (P*), Vargas sidesteps commitment to (P) and its discredited claim that most people have the intuition that determinism precludes full moral responsibility. Of course, once Vargas has abandoned (P) he can perhaps no longer claim to be a folk incompatibilist; but he can still maintain that the majority of us have demonstrably libertarian leanings, a claim that (P*) is intended to capture (and so we might label him a 'folk libertarian').

As far as the evidence for (P*) goes, Vargas is on stronger ground. Recalling Nichols and Knobe's experiment in which two Universes, Universe A and Universe B (a deterministic and an indeterministic universe, respectively), were described to participants who were then asked: 'Which of these universes do you think is most like ours?' According to Nichols and Knobe, nearly all participants (over 90%) considered Universe B (which was described as being determined in all aspects with the exception of human decisions, which, participants were told, were not completely caused by the past and thus did not have to happen the way they do) to be the universe most similar to our own.[28] It seems, then, that prediction (a) finds clear evidential backing in Nichols and Knobe's experiment.

As for prediction (b) set out in (P*), it is not clear whether participants in Nichols and Knobe's experiment were asked whether full moral responsibility is possible in the undetermined Universe B, since no data has been provided. In any case, the prediction that a majority of people believe that full moral responsibility is possible in our own universe is scarcely one that requires supporting evidence—the claim that, under certain conditions, an agent is fully morally responsible for their behaviour has the status of a truism. If this contention is correct, then we can infer not only that the vast majority of people think that full moral responsibility is compatible with existence in a Universe B-type world, but that we and our world provide evidence of that compatibility.[29]

[28] Nichols and Knobe (2007), p. 669.

[29] Indirect evidence that most people consider it possible to be fully morally responsible in our universe comes from the Roskies and Nichols (2008) experiment cited above. Since the majority of participants in their experiment believed full moral responsibility would be possible even if scientists

Before moving on from the prescriptive to the descriptive element of Vargas's moderate revisionism, there are two fundamental points to be gleaned from the discussion so far. The first is that most of us do indeed believe that we have both libertarian free will and full moral responsibility. In other words, the evidence from Nichols and Knobe validates prediction (P*). A second point, however, is that this evidence for what can be termed 'folk libertarianism' of the sort expressed by (P*) does not amount to evidence for folk incompatibilism as expressed by (P). In fact, there is every reason to think that (P)—the claim that the majority of people are categorical incompatibilists—is false, since, when presented with a this-world scenario in which determinism has been found to be true, the majority will choose to maintain their belief in full moral responsibility at the expense of their belief in libertarian free will.[30] This is an interesting result, and one that suggests, say Roskies and Nichols, that the intuition that we are morally responsible is what they call a "non-negotiable intuition."[3132] Since it remains a possibility that scientists might find empirical evidence for global determinism, it cannot therefore be claimed (contrary to what (P) predicts) that the majority of the folk are strict incompatibilists: if determinism turned out to be true, the vast majority would sooner concede the truth of compatibilism than abandon belief in moral responsibility.

In summary, then, we can say that the experimental results provide confirmation of (P*). Additionally, the above results suggest that the great majority of people would abandon their belief in indeterminism more readily than they would forsake belief in moral responsibility. The folk position on free will and moral responsibility thus appears to be curiously congruent with the one presented by van Inwagen, who has described his own libertarianism as involving a hierarchy of beliefs in which the

---

were to discover that our world is determined, it can be reasonably inferred that a still higher proportion would believe in the possibility of moral responsibility if it were conclusively proved that determinism is false.

[30] See both Roskies and Nichols (2008) and Nahmias *et al.* (2006).

[31] Roskies and Nichols (2008), p. 12.

[32] Another way of explaining folk intuitions regarding determinism and moral responsibility would be to say that the majority of people accept a "non-robust conditional," a term due to Frank Jackson from his (1991) monograph, *Conditionals*. A non-robust conditional is one that a person accepts, but would reject if there were sufficient evidence for the truth of the antecedent. In this instance, the antecedent is that determinism is true, while the consequent is that full moral responsibility is impossible. So, most people are willing to accept a conditional which states that the truth of determinism would entail the impossibility of full moral responsibility (and judgments on scenarios relating to other worlds bear this out, as the majority of people display incompatibilist intuitions in such cases); but once asked to suppose that the antecedent (i.e. determinism) is true of our world, most people are inclined to reject the conditional rather than accept its consequent (i.e. the impossibility of full moral responsibility).

reality of moral responsibility is held to be utterly foundational and incontrovertible, while the belief in indeterminism is strongly held yet tractable.[33] The final word on the issue of our folk intuitions is that Vargas is entitled to claim that the majority of people are folk libertarians, on the proviso that he understands that claim to be an endorsement of (P*) as opposed to (P). As for further intuition-testing experiments, it would be interesting to discover just to what extent the folk think it likely that we have libertarian free will. The results indicate that belief in indeterminism is negotiable in a way that the belief in full moral responsibility seems not to be for so many—so what sort of probability, it may be asked, would the folk assign to the thesis that determinism is true?[34]

5.3 Debating Vargas's Prescription

The descriptive element of Vargas's theory appears to be secure - so long, that is, as he is willing to be understood as endorsing (P*) instead of (P). The prescriptive element is still more crucial, however, as this outlines what kind of theory Vargas thinks we should be adopting. It is clear that Vargas wishes to prescribe a compatibilist solution to the problem of free will and moral responsibility. But this now raises the vexed question of how Vargas understands his own assertion that, on his theory, people can be deserving of praise and blame. The thought is that, if Vargas wishes to provide a consequentialist justification for maintaining our current responsibility norms (or at least for maintaining similar norms to those to which we currently adhere), then he is in effect conceding that the notion of ultimate desert—the kind of responsibility with which Galen Strawson, among others, is concerned—should be abandoned. Pereboom sets out to disentangle the two different senses of moral responsibility as follows:

---

[33] Van Inwagen (1983), pp. 219-21.

[34] The question about our folk intuitions regarding the likelihood of determinism is unaddressed by the studies examined here, in all of which it was simply stipulated in the various this- and other-worldly scenarios that either indeterminism or determinism was true. If more were known about how plausible people generally consider the possibility of determinism to be, we would have a truer sense of the extent to which people are inclined towards folk compatibilism. It might turn out, for example, that most people believe there is a vanishingly small probability that human decisions in this world are determined—in which case the results of Roskies and Nichols' scenarios, in which they stipulated that determinism was true of our world, would offer far less support for folk compatibilism than might at first glance be presumed.

I am very much open to the view that the question: "Are we sometimes morally responsible for our actions?" as posed in ordinary language, needs to be disambiguated. If it is specified that moral responsibility in the "legitimately called to moral improvement" sense is meant, then the answer is "yes." This answer would quite obviously not be inconsistent with hard incompatibilism, nor to incompatibilism more generally, as these notions function in the debate. For what is at issue is whether moral responsibility in the "basic desert" sense is compatible with determinism (and with the relevant sorts of indeterminism).[35]

Recall how Vargas asserts that holding one another morally responsible for our good and evil deeds is (at least partly) justified by the fact that doing so will "tend to contribute to our better perceiving and appropriately responding to moral considerations."[36] Pereboom agrees that we all are morally responsible in *this* sense—in the sense that we can be "legitimately called to moral improvement"; but he suggests that this leaves open the question of whether we can be morally responsible in the "basic desert" sense, a question which he himself answers in the negative. So where does Vargas stand on the question of whether we can be ultimately responsible in the way that Pereboom denies is possible?

Vargas replies that this second notion of moral responsibility—i.e. moral responsibility in the "basic desert" sense—is "exactly the sense of moral responsibility with which we should be concerned," and with which he too is concerned.[37] While some degree of conceptual revision away from our libertarian commitments regarding this notion of moral responsibility may be required, he concedes, it is not his intention (in contrast with Pereboom) to eliminate the notion of basic desert altogether. People are deserving of praise and blame when they are responsible agents who have violated the norms of the responsibility system, claims Vargas, and since there are many instances of responsible agents violating the norms of the responsibility system there are correspondingly many instances of agents being responsible in the basic desert sense.

Vargas's statement of his own intent is clear: he does not wish to dispense with the notion that we are sometimes morally responsible in the basic desert sense. On this point he is at odds with hard determinists such as Pereboom (who sees the sourcehood condition as a requirement if we are to maintain the notion of basic

[35] Fischer *et al.* (2007), p. 200.
[36] *Ibid.*, p. 156.
[37] *Ibid.*, p. 210.

desert), but in agreement with compatibilists such as Fischer and Ravizza (who deny that the sourcehood condition must be met). While Vargas himself declares that he does not wish to dispense with the notion of basic desert, the question remains: does Vargas's theory of revisionism entitle him to retain this notion?

It is understandable that, in his discussions with Vargas, Pereboom should seek to clarify the distinction between two different notions of moral responsibility—as a legitimate call to self-improvement on the one hand, and in a basic desert sense on the other. After all, Vargas's emphasis on the centrality of utilitarian considerations might reasonably lead a person to believe that his concern lies solely with the notion of moral responsibility as a legitimate call to self-improvement, and not at all with the notion of basic desert. It is puzzling that this is not the case, too, since Vargas evidently sees himself as offering up a justificatory story that is distinctively different from his compatibilist contemporaries. Abandoning the notion of basic desert would certainly be a way for Vargas to mark his theory out as being set apart from these more conventional compatibilists, and would burnish his credentials as a fearless revisionist in the face of what might be characterised as the unthinking orthodoxy.

Perhaps, however, dispensing with basic desert is one revision too far for Vargas, and so, like Fischer and Ravizza, he affirms instead that "people can deserve praise and blame when they are responsible agents who have violated the norms of the responsibility system."[38] As for the origination condition, if the belief persists that this must be met in order for praise and blame to be deserved, then this must be revised away just as the alternative possibilities requirement and agent causal beliefs were subject to revision.[39]

Unfortunately, Vargas adduces no reasons for believing that his brand of revisionism entitles him to the notion of basic desert. His compatibilist prescription places him in the exact same predicament that faces all compatibilists, and no fresh solution is proffered. Of course, for those who side with compatibilists in thinking that the origination condition can be dispensed with, Vargas's revisionism does indeed

[38] *Ibid.*, p. 211.

[39] Vargas (in Fischer *et al.* (2007), p. 215) writes: "On the variety of revisionism I favor, we should revise our commonsense construal of the alternative possibilities requirement, any agent causation elements in our thinking, and, if we have them, any incompatibilist conception of a sourcehood requirement."

present a viable alternative. However, in the present dialectical situation, conventional compatibilist responses are considered unsatisfactory since they fail to account for the intuition that the absence of origination precludes the possibility of what we have been referring to variously as 'basic desert' and 'ultimate moral responsibility' (in Pereboom's and Galen Strawson's terminology, respectively). Vargas offers no novel solution to the problem on this point, and so, despite the ingenuity of his theory, it fails to offer a wholesale solution to our predicament. Therefore, despite Vargas's insistence to the contrary, I find myself in agreement with Pereboom's appraisal that: "[I]f we revised our notion of free will to a compatibilist one, we would also need to revise our notion of moral responsibility so that the "basic desert" sense is eliminated."[40] Since the prescriptive element of Vargas's theory is essentially compatibilist, he is unavoidably committed to the elimination of the notion of basic desert, irrespective of the fact that he refuses to concede that this is the case.

Before moving on to examine our second 'mixed view', it should be said that while Vargas's theory does not provide us with the complete solution we would like, it does offer us the intriguing possibility of exploring whether a consequentialist justification of our moral responsibility norms might help resolve our compatibilism/incompatibilism dilemma. In other words, while Vargas himself does not wish to revise our concept of moral responsibility in such a way that the notion of basic desert is eliminated, there is no reason for us not do so – and at the same time, there is no reason for us not to retain a purely consequentialist justification of our moral responsibility norms. In this way, we would be drawing inspiration from Vargas by retaining what is distinctive and different about his approach to moral responsibility norms, and indeed taking his revisionary spirit still further by (further) redefining what it means to be morally responsible. This possibility will be explored later.

[40] Fischer *et al.* (2007), p. 199.

5.4 Double's Free Will Subjectivism

The mark of a 'mixed view' is that it favours neither the standard compatibilist nor the standard incompatibilist positions on free will, instead combining elements of both in an attempt to offer a fresh and illuminating new perspective. Our next 'mixed view' has perhaps an even greater claim to impartiality between the arguments of compatibilists and incompatibilists than Vargas's, since, while Vargas finds in compatibilism's favour by affirming the truth of the claim that free will (of the sort required for moral responsibility) is compatible with determinism, the view now under consideration shows no such partiality. Instead, it simply denies that the claim that moral responsibility is compatible with determinism possesses any truth value whatsoever. As such, neither compatibilists nor incompatibilists are right to think that free will and determinism are compatible. (Or perhaps both are right, if we prefer to think charitably about it.)[41]

The present view belongs to Richard Double, and he terms it "Free Will Subjectivism." As Double explains, free will subjectivism is the position of denying that "judgments concerning moral freedom provided by the lower-level free will theorists can be objectively true."[42] The lower-level free will theorists Double refers to here all hold to one of the four following positions: traditional compatibilism; traditional incompatibilism; the no-free-will-either-way theory; and the free-will-either-way theory. Double's designation of these theories as 'lower-level' is not made with any pejorative intent—they are lower-level merely in the sense that they are concerned with how to approach questions of moral freedom. By contrast, 'meta-level' theories are essentially concerned with the question of whether we should be subjectivists or objectivists concerning our lower-level judgments. On this question, Double insists we must be subjectivists about our free will judgments.

---

[41] I am not merely being flippant (although I am being a bit flippant) in suggesting that Double's theory might allow us to vacillate between thinking that compatibilists and incompatibilists might both be right, and thinking that they might both be wrong: Double himself seems inclined to think this way too. For example, when talking about conflicting lower-level theories, Double (2004) claims that both can be "equally true" (p. 415); yet he also denies that lower-level theories can ever be "objectively true" (p. 413). The key to this apparent paradox, I think, is to realise that conflicting lower-level theories can both be true, but only subjectively, while no lower-level theories—whether conflicting or not—can be objectively true.

[42] Double (2004), p. 413.

As Double makes clear, this distinction—one between lower-level and meta-level theories—echoes a parallel distinction in the field of ethics. Here too we see a debate at the meta-level over whether moral judgments can be objectively true: on one side of the debate, meta-ethical objectivists insist that moral judgments possess objective truth value, while meta-ethical subjectivists deny this.[43] Both objectivists and subjectivists engage in lower-level, normative ethical debates, attesting to the fact that subjectivists are capable of arguing just as vociferously for their normative ethical theory as objectivists.[44]

The terms "objectivism" and "subjectivism" get used in a great variety of ways, of course, so it would be wise to clarify just how Double intends them to be understood. In Double's own words, these terms are primarily intended to "refer to metaphysical theses about the 'location' or ontological dependence of entities."[45] To elucidate, Double asks us to consider the case of peas. We can say of peas that they exist objectively, or in their own right, whereas the taste of peas exists subjectively in persons who like or dislike their flavour. So, by analogy, just as it would be wrong to claim as an objective fact that peas are disgusting, it would be wrong to claim as an objective fact that an agent is morally responsible in any given situation. As a free-will subjectivist, Double thus believes that there can be no objective fact of the matter about which of the four lower-level theories of free will is true: instead, judgments affirming or denying moral freedom are "dependent on the feelings and attitudes of the persons who think about such things."[46]

Given that Double believes there is no objective fact of the matter about which lower-level theory is true, what implications does this have for his views on moral responsibility? In particular, does Double's belief that judgements of moral responsibility lack true value lead him to disavow any right to an opinion on such judgements? No, he does not abandon all opinions in the light of his commitment to free will subjectivism, but rather feels at liberty to pick and choose among the lower-

[43] Double is himself a meta-ethical subjectivist, and although he does not explore the issue, it seems hard to imagine that anyone might be an objectivist in one domain and subjectivist in the other. If this is right, then it is all the more strange that the possibility of free will subjectivism has to date remained so little explored, as there are no shortage of philosophers happy to proclaim themselves subjectivists in the realm of meta-ethics.
[44] See Hume (1965), Mackie (1977), and Smart (1973) for examples of metaethical subjectivists who express strong opinions on normative ethics.
[45] Double (2004), p. 412.
[46] *Ibid.*, p. 413.

level theories as his feelings and attitudes dictate: "Accepting free will subjectivism leads *me* to a fairly mixed acceptance of the lower-level theories" writes Double, "inasmuch as all of the theories seem attractive to me in many cases."[47] The key point to note, of course, is that he chooses among the lower-level theories *on the basis of his feelings and attitudes*, and not on the basis that there can ever be an objectively correct response when it comes to judgments of an agent's free will and moral responsibility.

What advantages are there in free will subjectivism? Double argues that his theory is liberating, in that it allows us greater resources when addressing questions of moral responsibility, reward, and punishment. We can opt for any lower-level theory we want, which means that, unlike the free will objectivist, we need never abandon our most strongly-held moral intuitions even when these are in lower-level conflict. On this point, Double's free will subjectivism provides an interesting contrast to Vargas's revisionism: Double considers it an advantage of his theory that he need not revise any of his moral intuitions for the sake of what he sees as a misplaced belief in the importance of theoretical consistency. In fact, Double goes so far as to claim that it is "*morally* better to keep our philosophical moral judgments as close as possible to our moral intuitions, because our best-considered moral feelings and attitudes constitute morality."[48] So, far from adopting Vargas's position of endorsing revisions in our beliefs concerning moral responsibility, Double takes the opposing line that revision is best avoided if at all possible.[49]

A moral advantage of free will subjectivism, according to Double, is that adopting free will subjectivism might make a person less disposed to blaming and punishing. Double quotes Waller in this connection: "when it is understood that there is no deep objectivity in our notions of just desert, that will leave us much less inclined to punish, and more inclined to look for causes that can be corrected."[50] Nonetheless, Double insists that "as subjectivists we can still apply moral responsibility when we want it to apply," much in the same way that the normative ethicist can endorse a

[47] *Ibid*., p. 413.
[48] *Ibid.*, p. 416.
[49] It is perhaps a little ironic, therefore, that proponents of the four lower-level theories would need to revise their beliefs regarding free will and moral responsibility very substantially were they to be persuaded that Double (2004: p. 413), the self-confessed "lone meta-level subjectivist," is right to spurn objectivism regarding moral responsibility judgments.
[50] In correspondence with Double (2004), p. 413.

deontological, consequentialist or virtue ethicist theory while remaining a meta-ethical subjectivist.[51]

Another perceived advantage of free will subjectivism—and one that is related to the fact that it allows us to avoid revising our strongly-held intuitions—is that holding to it represents a prudent strategy. Even if it turns out that free will subjectivism is wrong and that there is after all an objective fact of the matter concerning which lower-level theory is true, the adoption of meta-level subjectivism allows us to hedge our bets, as it were, and so avoid the risk of choosing the wrong lower-level theory. Of course, if it turns out that meta-level subjectivism is wrong, some of the moral intuitions to which we hold as spread-betting meta-ethical subjectivists will of necessity be wrong; but this is no reason to favour meta-ethical objectivism, as Double argues:

> [E]ven if we stipulate that meta-level objectivism is true, we still would be no closer than we are now to knowing *which* lower-level theory is the true one. If we cannot know which objectivist theory is correct (and the history of the free will debate suggests this is not likely to change), then we cannot know which moral intuitions to surrender.[52]

In other words, even if free will subjectivism is wrong and some particular lower-level theory does represent the objective truth, there is only a 25% chance that we will choose to endorse the correct lower-level theory out of the four on offer (assuming of course that each is equally likely to be correct). Better, then, that we remain free will subjectivists, safe in the knowledge that at least some of our intuitions regarding moral responsibility will be objectively correct should it turn out that free will subjectivism is false.

A final question regarding free will subjectivism: what does Double's theory have to say about the crucial issue of origination? First, Double acknowledges the truth of Honderich's observation that "if we have life-hopes for ourselves as undetermined "originators" of our choices, these hopes will be rationally unavailable to us if we accept determinism at the macro-level."[53] Double thus accepts that determinism rules out origination, and also acknowledges that it is part of (many of) our life-hopes that we should be undetermined originators of our choices. However, he goes on to

[51] Double (2004), p. 413.
[52] *Ibid.*, p. 416.
[53] *Ibid.*, p. 418.

suggest that a strength of free will subjectivism is that it allows us, when we "elect to wear our lower-level compatibilist spectacles," to divorce judgments of moral responsibility from feelings about life-hopes regarding origination.[54] Even-handed to a fault, Double concludes that he can see why someone would want to make the connection between moral responsibility and the need for origination; and he can also see why one would not. And according to the dictates of free will subjectivism, there is no compulsion for anyone either to affirm the necessity of origination for moral responsibility, nor, conversely, to deny it. Since there can be no objective fact on the matter of origination, one is at liberty to do either, or neither, or both, without fear of contradiction.

5.5 Why Free Will Subjectivism is Inadequate

To cut to the chase, I believe that Double is wrong to hold to the doctrine of free will subjectivism. Moreover, he is wrong to hold to free will subjectivism because the theory is incorrect. While it might seem obvious that we should reject Double's theory if it is deemed incorrect, it should be remembered that Double argues his theory should be adopted even if it is suspected of being false. This line of argument I shall dub his 'prudential justification' for free will subjectivism. Therefore, when I reject the theory on the grounds of judging that it is likely to be false, I am declaring that the 'prudential justification' for being a free will subjectivist is not strong enough to counterbalance consideration of what I consider the likely falsity of the theory. In this section, I will explain why I think that free will objectivism is true, a task that will involve responding to Double's points in favour of his own theory as set out in the previous section. Nevertheless, in spite of my rejection of free will subjectivism, I shall argue that Double's 'prudential justification' provides us with a valuable lesson about the importance of having the humility to hedge one's bets when the situation requires.

We have seen how Double declared it an advantage of his theory that it allows us to avoid revising any of our beliefs. Instead, we are free to maintain whatever beliefs we wish about moral responsibility without fear of being forced to give them up on

[54] *Ibid.*, p. 418.

account of any theory. This claim—that we are free to make any judgment whatsoever about any situation of possible moral culpability—really does grant us an incredible degree of freedom in our thinking on moral responsibility; but of course it also leads to an obvious criticism, which is that free will subjectivism ultimately grants us too much freedom. With no objective truth to anchor us, both theoretical and practical problems emerge. On the issue of practical concerns Double assures us that, as subjectivists, we can "still apply moral responsibility when we want it to apply."[55] Nonetheless, despite this assurance, the sceptical question naturally arises: on what authority can it be applied if it is widely accepted that no judgment of moral responsibility can be objectively correct? Furthermore, what happens when others do not want moral responsibility to apply? Of course, it must be said that we encounter cases of conflicting beliefs over the proper assignment of moral responsibility all the time. The difference between these commonly encountered cases and those that would arise if belief in free will subjectivism were the norm, however, is that the former are held against a backdrop of shared belief in the existence of objective facts of the matter on the issue. Without a shared belief in objective truths concerning moral responsibility judgments, it is hard to see how we could even attempt to resolve conflicts of opinion on questions of moral responsibility.

Who has the authority to impose their subjective judgments, and how conflicts in judgments are to be resolved are important concerns; but they are also essentially practical concerns, in that to raise them is to raise doubts about the desirability of a widespread belief in free will subjectivism. Such practical concerns lead naturally, however, to deeper questions about the very intelligibility of free will subjectivism. The main worry, I would argue, runs as follows: if free will subjectivism is true, then it would not be legitimate to offer reasons in support of one's subjectively-held judgments of moral culpability. Of course, in common with all people, philosophers and non-philosophers alike, Double cannot refrain from offering reasons and justifications for his moral judgments, and these, in consequence, belie his insistence in the truth of free will subjectivism.

Why, if free will subjectivism were true, would it not be legitimate to offer reasons in support of one's subjectively held judgments of moral culpability? Because,

[55] *Ibid.*, p. 413.

according to free will subjectivism, judgments affirming moral freedom and responsibility are supposed to have their source in one's feelings and attitudes, and these feelings and attitudes need not—and indeed cannot—be justified by reference to any reasons. That is, if judgments of moral responsibility are at root merely subjective feelings and attitudes, then these feelings and attitudes not only require no explanation, but they must not be amenable to explanation, since if they were it would imply that they had some objective basis that others could be compelled to recognise. Not surprisingly, of course, Double makes frequent reference to his reasons for holding certain supposedly subjective moral judgments, and, more problematically still, he sees no problem with having reasons for holding conflicting opinions at one and the same time.[56] The basic thought against Double here is this: to offer reasons for your opinions regarding moral responsibility is to tacitly accept that it is open to the objective evaluation of others, on the basis of those reasons adduced for holding them, to determine whether or not these opinions are justified.

I contend that Double's appeals to reasons for his holding certain moral responsibility judgments betrays an unconscious objectivism in his thinking, in spite of his stated allegiance to free will subjectivism. We can also note Double's tendency to stray into objectivism when criticising lower-level, competing theories. According to the logic of his free will subjectivism, I would argue that Double is considerably constrained in what he is at liberty to say in response to other theories. As I have already claimed, he cannot offer reasons for why his judgments of moral responsibility are superior to someone else's, since his judgments are supposed to be mere feelings and attitudes. By the same token, he certainly cannot claim that his theory is right and another one wrong, since this is even more transparently the language of objectivism. In fact, he must refrain from offering any kind of critique that presupposes or otherwise implies that there is some objective standard to which his philosophical adversary should be receptive. Yet this is precisely what he does when arguing against Pereboom. In response to Pereboom's assertion that a quarantine-rehabilitation rationale for incarcerating criminals is justified on the basis

[56] Double (2004: p. 415) says, for example, that Martin Luther was praiseworthy "for his reason-sensitive stand against the Church," yet also not praiseworthy "inasmuch as his actions were caused by heredity, by environment, and, if libertarianism is the case, partly by chance"; he also suggests that subjectivists should deny that Stalin is morally responsible for the Purges at the lower level (since according to hard determinism and no-free-will-either-way-theory, no one is morally free); while at the same time holding that Stalin (rather than Trotsky) *is* morally responsible at the lower level "when we wish to single out the correct person for assigning moral responsibility."

of his no-free-will-either-way theory, Double imagines a scenario in which, according to this quarantine-rehabilitation rationale, we should incarcerate criminal B for a crime in fact committed by A, due to B's comparative malleability and the likelihood of him falling into a life of crime unless checked. Commenting on this scenario, Double protests that it "seems unfair" to quarantine B.[57]

While I can sympathise with Double's sentiment here (incarceration for a crime one has not committed is, after all, manifestly unjust), I question whether he is entitled to disapprove in such terms—or, at least, I question whether his disapproval, coming as it does from a free will subjectivist, carries any weight. Since Double is a subjectivist, I struggle to understand what he means to convey in asserting that the no-free-will-either-way has "unfair" consequences, such as the incarceration of innocents. Specifically, I worry that, in declaring the incarceration of an innocent "unfair," he is appealing to some objective fact of the matter, which is something he is not entitled to do, of course. The dilemma as I see it, then, is as follows: Double is either appealing to an objective notion of fairness to which he is unentitled; or else he is appealing to a subjective notion of fairness, and thus Pereboom—whose own notion of fairness cannot then be deemed false nor its value be gainsaid—need pay no heed whatsoever to Double's criticisms.

I have argued that Double's appeal to reasons in support of his own ascriptions of moral responsibility, and his choice of language when critiquing lower-level theories, both unwittingly betoken a belief in an objective standard as regards moral responsibility judgments. While it is possible that Double could explain why this is not the case, these considerations certainly threaten to impugn his theory of free will subjectivism. But perhaps the most compelling consideration of all in favour of rejecting Double's theory is that it fails to do justice to our intuitions—an ironic fact, given that he argues free will subjectivism allows us to cleave to whatever moral responsibility intuitions we might already have. In fact, the one big intuition we are obliged to surrender is the belief that our moral judgments have a truth value. To illustrate this with reference to the issue of origination, we have seen how Double understands why someone would want to make the connection between moral responsibility and the need for origination, and he can also understand why one

[57] Double (2004), p. 417.

would not. And ultimately it is a mere a matter of opinion, as far as Double is concerned, whether and on what occasions one decides to affirm the necessity of origination for moral responsibility. This so fails to capture my own intuition—that, on the contrary, when we debate issues such as the necessity or otherwise of origination, we are surely arguing over objective facts rather than mere opinions—that I can only conclude that Double's theory must be rejected out of hand. The inducement that one is allowed to maintain all of one's lower-level intuitions in becoming a free will subjectivist are, to my mind, far outweighed by the heavy price that must be paid in the form of the necessary abandonment of belief in objective truths regarding moral responsibility.

To finish discussion of free will subjectivism, let me say a few more words about Double's 'prudential justification' for the theory, the gist of which is that it is prudent to be a free will subjectivist even if the theory should turn out to be false, since we are likely to endorse the wrong lower-level theory anyway. While feeling very much at liberty to declare that free will subjectivism is false beyond reasonable doubts, the question remains: is there nonetheless a case for being a free will subjectivist in the light of the 'prudential justification'?

Again, the answer is 'no,' quite simply because the overwhelming likelihood of the falsity of free will subjectivism provides all the warrant needed for rejecting it. That rejection comes with a caveat, however, which is that we have cause for circumspection regarding our ability to correctly discern the objective facts of the matter when it comes to moral responsibility. This becomes evident when one reflects on the reality that there remain large and seemingly intractable disagreements between lower-level theorists, and thus to opt for the wholesale endorsement of any one of these theories is to risk backing the wrong horse, so to speak. While this concern does not amount to a reason for adopting subjectivism, it may well provide the motivation to favour sticking to one's intuitions as opposed to forcing one's intuitions to conform to whichever lower-level theory is thought to have the best credentials. In matters of prudence, then, if not on any other matter, Double's free will subjectivism has something to teach us.

6

Smilansky's Mixed View

With Vargas and Double's theories having been surveyed, this chapter brings us to consideration of our final mixed view. This is articulated by Smilansky, who believes that, in a world in which libertarian free will does not exist, compatibilism and incompatibilism nonetheless each fail to capture adequately what he considers to be our well-founded intuitions on free will and moral responsibility. To remedy this perceived failure in the theories of compatiblism and incompatibilism, Smilansky offers two distinctive proposals. The first is that we should embrace what he terms the 'fundamental dualism.' In a nutshell, this sees Smilansky arguing that, while neither compatibilism nor incompatibilism is adequate on its own, aspects of both are indispensable, and we must therefore abandon what he refers to as the 'assumption of monism,' according to which either compatibilism or incompatibilism must be affirmed wholesale while the other must be rejected in its entirety. Smilansky's second proposal is that we embrace 'illusionism' with regards to free will. According to Smilansky, humanity as a whole "is fortunately deceived on the free will issue, and this seems to be a condition of civilized morality and personal value."[1] It is not that we should be inducing illusory beliefs, says Smilansky, and still less that we might be able to live with beliefs that we fully realise are illusory: rather, the contention is that illusory beliefs are already in place in many of us, and that (generally speaking) these play a positive role.

Each of these proposals will be unpacked in greater detail in the course of this chapter, and their merits and deficiencies will be weighed. Once Smilansky's theory has been presented and evaluated, I shall reflect on what it has to offer us in terms of helping to create a credible theory of moral responsibility.

[1] Smilansky (2002), p. 500.

6.1 Smilansky's Fundamental Dualism

While Smilansky's theory is comprised of two distinctive proposals which he seeks to persuade us to adopt, it is perhaps the first of these—fundamental dualism—that most clearly provides the warrant for labelling his theory a mixed view. The proposal of a fundamental dualism is offered as an alternative to what Smilansky characterises as the prevailing orthodoxy, the assumption of monism, according to which one must affirm compatibilism or hard determinism.[2] In fact, Smilansky demurs, "there is no conceptual basis for thinking that the Assumption of Monism is necessary."[3] This denial of the necessity of the assumption of monism does not amount to an affirmation of the logical consistency of compatibilism and hard determinism—on the contrary, Smilansky acknowledges their logical inconsistency. Rather, it points to the possibility of holding a "mixed, intermediate position that is not fully consistent with either."[4]

Before we go on to see what such a mixed, intermediate position might look like, a few words should be said on why Smilansky considers both compatibilism and hard determinism insufficient. Beginning with compatibilism, it is evident that Smilansky is moved by considerations of origination, control, and luck—the same issues that motivate the previously discussed arguments of van Inwagen and Galen Strawson—to declare that ultimate responsibility is impossible. Consider the following passage:

> If people lack libertarian free will, their identity and actions flow from circumstances beyond their control. To a certain extent, people can change their character, but that which does or does not change remains itself a result of something. There is always a situation in which the self-creating person could not have created herself but was just what she was, as it were, "given." Being the sort of person one is and having the desires and beliefs one has, are ultimately something one cannot control, which cannot be one's fault; it is one's luck. And one's life, and everything one does, is an unfolding of this.[5]

From what Smilansky says here, it seems he is persuaded of the validity of the arguments for hard determinism discussed in Chapter 4; in fact, however, his rejection of compatibilism does not lead him to unqualified acceptance of hard

---

[2] As for libertarianism, Smilansky rejects this on the grounds that indeterminism cannot contribute anything to moral responsibility.
[3] Smilansky (2002), p. 491.
[4] *Ibid.*, p. 491.
[5] *Ibid.*, pp. 492-3.

determinism.[6] Just as compatibilism is deemed insufficient, so too is hard determinism, on account of the fact that we must retain certain important compatibilist distinctions if we are to do justice to morally required forms of life. Smilansky believes that such distinctions as compatibilists are wont to make often have crucial, non-consequentialist, ethical significance. To support his contention that compatibilist distinctions are sometimes unjustifiably overlooked by hard determinists, Smilansky offers the following reflection:

> [T]he kleptomaniac and the alcoholic differ from the common thief and common drinker in the deficiency of their capacity for local reflective control over their actions. Here everyone should agree. But the point worth adding is that such differences are often morally significant.[7]

Smilansky is arguing that there are sometimes morally salient differences between two cases (such as between a kleptomaniac and a common thief) which the compatibilist is able to recognise but which the hard determinist would ignore. In such cases, the intuition is that it is simply not fair to regard both agents as equally lacking in moral responsibility. Furthermore, we as agents do not *want* to live in a society in which its members are morally excused for their bad deeds where no mitigating circumstances exist:

> We want to be members of a Community of Responsibility where our choices will determine the moral attitude we receive, with the accompanying possibility of being morally excused when our actions are not within our reflective control […] We have to *enable* people to live as responsible beings in the Community of Responsibility, to live lives based largely on their choices, to note and give them *credit* for their good actions, and to take account of situations in which they *lacked* the abilities, capacities, and opportunities to choose freely and are therefore not responsible in the compatibilist sense.[8]

---

[6] In fact, Smilansky is not even committed to the acceptance of determinism, let alone hard determinism. That is, whereas I have argued that we must accept the truth of determinism on the basis of the *PSR*, Smilanksy considers it an open question whether determinism is true. However, *if* determinism proves to be true, then Smilansky would, I think, assent to the conclusion of van Inwagen's Consequence Argument. And if determinism proves to be false, then Galen Strawson's Argument for the Impossibility of Moral Responsibility would in any case suffice to convince Smilansky that ultimate responsibility is unattainable. Given his ambivalence regarding the truth of determinism, Smilansky might be more accurately characterised here as displaying sympathy for the doctrine of 'impossibilism'—and, insofar as I accept Galen Strawson's arguments, I share his impossibilist sympathies. Nevertheless, I shall continue to frame the debate as one between *hard determinism* and compatibilism, for two reasons: first, because this is how Smilansky himself frames it; and second, because I have already put forward the case that determinism is true.

[7] Smilanksy (2002), pp. 493-4.

[8] *Ibid.*, p. 495.

A partial condemnation of both compatibilism and hard determinism thus follows. On the one hand, Smilansky rebukes compatibilists on the grounds that any factor for which a person is appreciated, praised, or even loved is ultimately a matter of luck. "That compatibilists are indifferent to such ultimate arbitrariness, shallowness, and injustice," says Smilansky, "is morally outrageous."[9] On the other hand, he also fulminates against hard determinists for their failure to recognise the difference between cases such as that of the common thief and the kleptomaniac, and for being blind to the moral inadequacy of social institutions that would fail to take account of this difference. "That hard determinists are indifferent to such distinctions and ethical imperatives," Smilansky once more declares, "is morally outrageous."[10]

Smilansky's arguments for the partial validity of compatibilism and hard determinism (or, to put the same point in more negative terms, the partial inadequacy of both) sets the stage for his first proposal, which is the case for a fundamental dualism on the question of the compatibility of free will and determinism. There are aspects of each position—compatibilist and hard determinist—that the other cannot plausibly deny, and therefore any "monistic" position will be inadequate. In other words, we cannot completely disregard the hard determinist insight (as the compatibilist would have us do) that even vicious and compatibilistically free criminals are in some sense victims of circumstance. Neither can it be ignored that favourable compatibilist assessments of persons are necessarily shallow, since any such assessments ultimately depend on factors not under the person's control. Equally, however, there can be no disregarding the compatibilist insight that there are legitimate moral distinctions to be made between, say, the alcoholic and the common drinker, or the kleptomaniac and the common thief. The kleptomaniac, for instance, is simply not in a condition for membership in the kind of 'community of responsibility' to which most people, including the common thief, can—and should want to—belong.

How should we understand Smilansky's notion of a community of responsibility? The first thing to say is that such a community must be built on compatibilist distinctions, since the hard determinist perspective simply denies the reality of responsibility. Being built on compatibilist distinctions, a community of

[9] *Ibid.*, p. 497.
[10] *Ibid.*, p. 496.

responsibility would in one sense be unjust, says Smilansky, by virtue of the fact that it involves holding one another responsible even though our actions are not on the ultimate level up to us. In another sense, however, the community of responsibility *is* just, as it allows us to maintain the kinds of distinctions—between kleptomaniacs and common thieves, for instance—we already commonly make when deciding whether an agent's action was 'up to' them. To fail to create a community of responsibility, argues Smilansky, would be to fail to show proper respect for persons, because showing proper respect for persons requires the establishment of some kind of non-arbitrary moral order. Since no satisfactory moral order can be established merely on the basis of hard determinism (so Smilansky claims) we must therefore work with compatibilist notions of fault and moral responsibility. We can and should acknowledge that working in this way has a moral price in terms of unfairness and injustice when viewed from the hard determinist perspective, concedes Smilansky, but this does not negate the fact that there is a frequent need to think along compatibilist lines in order to maintain a community of responsibility.

In brief, the core insight of Smilansky's fundamental dualism is that, instead of embracing the mistaken assumption of monism (according to which either compatibilism or incompatibilism is wholly correct) we "need to try out ways of combining them."[11] Although Smilansky offers suggestions as to how they might be combined he allows that his particular suggestions need not be followed zealously, since his main aim is merely to illustrate the possibility of working within a dualistic framework. Whether Smilansky's idea of embracing fundamental dualism has merit will be considered following an exposition of his second proposal: 'illusionism.'

6.2 Smilansky's Illusionism

Smilansky's illusionism is comprised of two claims: the first of these is that the majority of people have illusory beliefs regarding free will; the second is that this situation is largely for the best. We will examine each of these claims in turn.

According to Smilansky, we—or at least the majority of us—have illusory beliefs regarding free will. So what does the notion of "illusion" mean to Smilansky, why

[11] *Ibid.*, p. 497.

do we need it, and exactly what illusory beliefs does he identify? In clarification of the first question, Smilansky tells us that the sense of illusion he is interested in combines the falsity of a belief with "some motivated role in forming and maintaining that belief—as in standard cases of wishful thinking or self-deception."[12] There is no need to determine the current level of illusion concerning free will, argues Smilansky, although he is in no doubt that it has a large role to play in the free will debate.

Why does Smilansky believe certain illusions regarding free will are necessary? The necessity of illusion arises, we are told, from two features of the free will problem: first, it arises indirectly as a result of the fundamental dualism—that is, it arises from the partial and varying validity of compatibilism and hard determinism. Smilansky terms the circumstances giving rise to illusion here the 'dissonance problem.' Second, the need for illusion flows directly and more deeply from the absence of libertarian free will, which would otherwise provide grounding for some of our beliefs on moral responsibility. This is labelled the 'insufficiency problem.' We shall examine each of these in turn.

If both sides of the fundamental dualism are acknowledged, warns Smilansky, then there is a risk that either compatibilism or hard determinism will be rejected in its entirety. This, in essence, is the dissonance problem. Smilanksy suggests various motivations that might lead someone to abandon one or other perspective: they may be troubled by the very existence of the dilemma, or else troubled by the fear of inconsistency or irrationality. Alternatively, perhaps one or other of the ultimate and compatibilist insights might encourage the rejection of one half of the fundamental dualism—for example, says Smilansky, "the idea that no ultimate-level distinction can be made might cause some people to neglect ultimate criteria, while causing others to discount all criteria."[13]

Whichever side of the fundamental dualism is abandoned on account of the dissonance problem, argues Smilansky, the result will be problematic. If the compatibilist perspective is rejected, then we can expect this to give rise to particular harms. The first of these harms is the *Present Danger of the Future Retrospective*

[12] *Ibid.*, p. 500.
[13] Smilansky (2000), p. 159.

Excuse, which Smilansky summarises as follows: "people ought not to be fully aware of the ultimate inevitability of what they have done, for this will affect the way in which they hold themselves responsible."[14] That is to say, if people are aware of the hard determinist insight that their actions are not ultimately under their control, the knowledge that this excuse for past behaviours will be open to them in the future is likely to encourage bad behaviour in the present. Another potential harm Smilansky cites from rejecting compatibilism is the *Danger of Worthlessness*. If it is believed that everything that happens is merely the unfolding of the given, then the worth of what one achieves as an individual might seem to be diminished. A final harm associated with rejecting compatibilism as a result of the dissonance problem is the *Danger of Retrospective Dissociation*. From the hard determinist perspective, an individual can easily appear as a mere vehicle, whose decisions, feelings, thoughts, and so forth are simply phenomena. Since having genuine feelings of responsibility are crucial to being responsible selves, argues Smilansky, we cannot remain sanguine about the dangers of adopting a stance of retrospective dissociation.

A rejection of the ultimate perspective does not pose the same risk in terms of potential harms that rejection of the compatibilist perspective poses. However, ultimate perspective rejection is equally unacceptable to Smilansky, for two reasons. For one thing, the ultimate perspective contains valid insights and as such cannot be wholly dispensed with any more than can the compatibilist persepctive. Additionally, ultimate-level conclusions are derived from a concern with such features as the source of—and control over—actions, issues which exercise compatibilists every bit as much as hard determinists. It is therefore not possible for compatibilists to discount hard determinists' concerns with these issues without, in so doing, diluting the power of their own theory.[15]

Whereas the dissonance problem arises indirectly from the absence of libertarian free will (via the partial validity of both hard determinism and compatibilism), the insufficiency problem arises directly from it. In particular, it arises "from the fact that much of our personal reactive life and closely related values, conceptions, and practices have deep connections with the idea of libertarian free will."[16] The danger,

[14] *Ibid.*, p. 153.
[15] See Smilansky (2000), pp. 149-61 for a lengthier disquisition on the Dissonance Problem.
[16] Smilansky (2000), p. 162.

according to Smilansky, is that if the reality of the absence of libertarian free will were made clear, then these values, conceptions, and practices would be abandoned *despite* being partially justified on compatibilist grounds. We therefore need the illusion of libertarian free will to provide a safeguard against the danger of moral nihilism that arises from exclusively embracing the ultimate perspective. So, as Smilansky argues: "if libertarian assumptions *carry on their back* [compatibilist] distinctions, which would not be adhered to sufficiently without them, an illusion which defends these libertarian assumptions seems to be just what we need."[17]

It is clear, then, that the three dangers discussed above in relation to the rejection of compatibilism due to the dissonance problem—the *Present Danger of the Future Retrospective Excuse*, the *Danger of Worthlessness*, and the *Danger of Retrospective Dissociation*—are equally applicable in the case of the insufficiency problem: all three are examples of what can happen when one adopts the ultimate perspective to the exclusion of any other, which is just what Smilanksy fears might happen should people become aware of the lack of libertarian free will. As Smilansky is keen to assure us, he is not arguing that there is no justification for our moral distinctions once it has been discovered that libertarian free will is false—on the contrary, compatibilism provides at least a partial justification for our moral responsibility practices. The problem is that Smilansky doubts whether most people would find compatibilist distinctions sufficiently compelling to prevent them from defaulting to a hard determinist denial of free will and moral responsibility. Given these facts, it is necessary for the "fragile compatibilist-level plants […] to be defended from the chill of the ultimate perspective in the hothouse of illusion."[18]

In summary, illusion is necessary, and results from both the dissonance problem and the insufficiency problem. Regarding the dissonance problem, if the absence of libertarian free will is realised, then there is the danger that one or other side of the fundamental dualism (either compatibilism or hard determinism) will be abandoned wholesale. If compatibilist distinctions are abandoned, this might lead to a moral nihilism threatening (among other things) our ability to hold ourselves morally responsible, to feel a sense of self-worth in our achievements, and to feel proper remorse for moral failings. If, on the other hand, the dissonance problem leads to the

[17] *Ibid.*, p. 173.
[18] *Ibid.*, p. 173.

rejection of the ultimate perspective of hard determinism, no such deleterious practical consequences will follow; but this outcome is nonetheless to be regretted on account of hard determinism's partial validity, as well as the weakening effect such a rejection would have on compatibilism itself given shared compatibilist/hard determinist concerns with questions of sourcehood and control. Smilansky is pessimistic about the hopes of successfully reconciling the two perspectives of the fundamental dualism: "Whichever balance between the elements occurs, complex patterns of confusion and unwarranted dismissal of one or both perspectives are very likely to emerge."[19]

As for the insufficiency problem, the difficulty here is that the absence of libertarian free will renders us unable to justify the many beliefs, values and practices related to moral responsibility that depend on the truth of libertarian free will. Our capacity to adopt an ethical view of one another, to feel worthy of deep respect, perhaps even our ability to view one another as autonomous individuals, are all potentially under threat from an awareness of the absence of libertarian free will. While it is true that compatibilist distinctions offer us partial justification for maintaining many of our beliefs, values and practices, many will consider this too weak a foundation on which to base their beliefs on moral responsibility. Such people are therefore likely to adopt to their detriment an exclusively hard determinist perspective. Additionally, even if one accepts compatibilist distinctions, they remain a poor substitute for belief in libertarian free will and the concomitant assurance of the truth of origination and ultimate responsibility.

For the reasons discussed, Smilansky commends to us his illusionism, which calls for "a large measure of motivated obscurity regarding the objections to libertarian free will."[20] Since Smilansky considers that these illusory beliefs are already largely in place, his illusionism can be seen as much as anything as a call to philosophers to keep silent about the absence of libertarian free will for the sake of the common good.

[19] *Ibid.*, p. 288.
[20] Smilansky (2002), p. 501.

6.3 Assessing Smilansky's Mixed View

Should we accept Smilansky's proposals for a fundamental dualism and for illusionism? I would argue that neither can be adopted in quite the form that Smilansky presents them. In fact, insofar as illusionism calls for motivated obscurity regarding objections to libertarian free will, I contend it should not be adopted at all. As for fundamental dualism, the suggestion that we need to combine compatibilist and incompatibilist ways of thinking shows more promise, and as such I hope to make the case that some sort of fundamental dualism should be adopted.

6.3.1 Assessing Illusionism

Let us first discuss the merits of illusionism, however, and in particular Smilansky's claims that illusion with regards to the non-reality of libertarian free will is widespread, plays a largely positive role in our lives, and ought to be supported on account of its usefulness. Beginning with the claim that illusion is widespread, we can note the similarity of Smilansky's and Vargas's positions on this point. While Vargas never describes belief in libertarian free will as illusory, he would certainly agree with Smilansky's conviction that such belief is both the norm, and is false.[21] It was argued in Chapter 3 that libertarianism is false. The question of whether belief in libertarianism is the norm was discussed in relation to Vargas's revisionism, receiving a qualified 'yes'—qualified principally by the observation that belief in moral responsibility seemed to be stronger than belief in indeterminism. In view of these facts, I have no objection to what we might (following Vargas) term the 'diagnostic' aspect of Smilansky's illusionism: libertarianism is indeed the predominant view in our society. Furthermore, it is false. As far as these diagnostic criteria are concerned, then, Smilansky's illusionism passes muster.

When it comes to the 'prescriptive' element of Smilansky's theory, though, things are trickier and more contentious. At the heart of the matter, however, is a simple question: is Smilansky right to think that, on balance, it is preferable to seek to

[21] Where Vargas is at variance with Smilansky is in the latter's claim that there is often a motivated element to our mistaken adherence to libertarianism: i.e. that self-deception is at play. Whether this claim—that adherence to libertarianism is largely motivated by the urge to self-deceive—is borne out by the evidence will be questioned shortly.

conceal from the majority the truth regarding our lack of libertarian free will? Smilansky presents two reasons for thinking that the absence of libertarian free will is a secret best kept hidden, and these together form his case for a "motivated obscurity" on the subject. The first reason cited is that illusion plays a positive role in its capacity as (in Smilansky's poetic phrase) "the handmaiden of reality" in disguising the absence of libertarian free will.[22] In more prosaic terms, the claim is that compatibilist distinctions, although possessing their own non-illusory reality, nonetheless require libertarian illusion for their support. To the extent that illusion helps to support these compatibilist realities, we can view this not as mere falsehood but rather as a condition for the emergence of a valid morally necessary reality. So, features of moral life such as the acceptance of personal responsibility, adherence to the values and practices of the compatibilistically-grounded community of responsibility, and a sense of pride in having done the right thing in a difficult situation, are all made possible by a belief (albeit false) in the reality of libertarian free will. As to why a belief in libertarian free will should be a condition for supporting compatibilist realities, Smilansky answers (as we have seen) that a realisation of the absence of libertarian free will would likely lead many to abandon valid compatibilist distinctions, and would encourage them to embrace the moral nihilism of hard determinism instead.

Another reason given in favour of a motivated obscurity with regards to the falsity of libertarian free will is that illusion can have a value in and of itself. In fact, Smilansky goes so far as to say that illusion "not only helps to create and sustain independent reality, but is *in itself* a sort of 'reality,' simply by virtue of its existence."[23] That is, the fact that a belief is false does not negate the fact that the belief is experienced as true by the believer. Moreover, the effects of such illusory beliefs can sometimes be positive, and Smilansky asserts that belief in libertarian free will is one such example of an illusory belief capable of giving rise to positive effects.

Despite Smilansky's pro-illusion arguments, I remain unconvinced. According to his first reason, in the absence of the illusory belief in libertarian free will, many of us would abandon perfectly valid compatibilist distinctions in favour of the ultimate

[22] Smilansky (2000), p. 170.
[23] *Ibid.*, p. 170.

perspective of the hard determinist. Illusion (in the form of belief in libertarian free will) is thus supposed to be a condition for supporting reality (the partial validity of compatibilist distinctions). However, I question the assertion that, on becoming aware of the absence of libertarian free will, many people would abandon valid compatibilist distinctions in favour of the supposed moral nihilism of hard determinism. To claim, as Smilansky is effectively doing, that most non-philosophically trained people require belief in libertarianism as a crutch, without which they would be unable to appreciate the (partial) validity of compatibilist distinctions, is to fail to credit such people with the natural intelligence they surely have. Smilansky's illusionism is excessively paternalistic in this regard, in that the general mystification and self-deception he perceives and endorses in the majority is deemed unnecessary for the minority. In short, since Smilansky is able to bear the knowledge that libertarian free will is a chimera, why should anyone else find it such a problem? What's good for the goose, as they say, is good for the gander.

Moreover, Smilansky's fear that moral nihilism awaits those that are stripped of their illusions regarding libertarian free will is hard to reconcile with the evidence considered in the previous chapter on folk concepts of free will. Nichols and Knobe's study, for instance, suggests that belief in moral responsibility is much more resilient than Smilansky realises. In particular, the finding that, if the world were found to operate according to deterministic laws, the majority of people would still judge an agent in such a world to be morally responsible, provides evidence that Smilansky's fears of moral nihilism are unfounded.[24] In fact, the power of belief in moral responsibility can be evinced simply by reflecting on the extent to which such belief is susceptible to elimination on theoretical grounds. Put simply, far from thinking that theoretical arguments for abandoning libertarianism are likely to lead people to abandon belief in moral responsibility, we should instead be concerned with the much more plausible worry of people continuing to praise, blame, punish,

[24] Nichols and Knobe (2007: p. 670) found that, when presented with a scenario in which an agent behaves immorally, 72% of participants judged that the agent was fully morally responsible for their behaviour even once it was stipulated that they lived in a determined world.

and reward in situations in which these practices have been found to have no theoretical warrant.[25]

Smilansky's final reason in favour of illusionism, which is that illusion can have value in and of itself, cuts little ice. It is doubtless the case that false beliefs and self-deception are sometimes valuable features of our psychological make-up, a fact evidenced by studies such as that by Alloy and Abramson, who found that supposedly 'normal' subjects in fact had inflated, and less realistic, perceptions of their importance, reputation, locus of control and abilities than did depressed subjects.[26] However, while most of us would surely agree that depression is too large a price to pay for the benefit of having a more realistic perception of one's character and abilities, the likely consequences of knowing the truth regarding libertarian free will are, I submit, significantly less onerous, to the extent that they fail to support the case for motivated obscurity. As for the observation that the falsity of a belief does not negate the fact that it is experienced as true by the believer, my response is: so what? The assertion is true, but I fail to see how it lends support to the claim that these false beliefs should merit protection from the harsh reality.

To conclude, I contend that the supposed threat posed by the unmasking of the illusion of libertarian free will is in fact no threat at all. While I accept that an understanding of the absence of libertarian free will is likely to be a cause for regret, one must balance that against the inherent value of knowing the truth, even when one would prefer that reality were otherwise. Admittedly, that it is always preferable for false beliefs to be challenged and revealed might be too strong a claim—if a more realistic outlook could only be achieved at the cost of serious depression, for example, then this would plausibly too high a price to pay. Another factor to consider is hinted at in the dictum that "a little knowledge is a dangerous thing": a complete lack of understanding is sometimes preferable to partial understanding, and this might provide grounds for motivated obscurity in some instances. However, in this particular instance—namely, regarding the commonly-held and erroneous belief in libertarian free will—it is hard to see that these grounds apply. Rather, Smilansky's dire prognostications for societal wellbeing should the truth of the lack

[25] I see support for my assertion here in P.F. Strawson's (1962) insight that our propensity to praise, blame, reward, and punish is so deep-rooted that it is questionable whether theoretical considerations could make any appreciable difference to such habits of moral appraisal.
[26] Alloy and Abramson (1979).

of foundation to belief in libertarian free will become widely known are, at the least, greatly overstated. Further, if Smilansky, being privy to this knowledge, finds that he is yet able to bear it without either falling into abject despondency or succumbing to nihilism, then why not expect the same of the man on the Clapham omnibus?

6.3.2 Assessing Fundamental Dualism

Leaving Smilansky's illusionism aside now and turning to his proposal for a fundamental dualism, I wish to argue that while this proposal must be adopted in some form, this need not—and in fact should not—be in the form in which Smilansky himself presents it. Fundamental dualism, it will be remembered, amounts to a repudiation of the assumption of monism, according to which one must accept either compatibilism and incompatibilism wholesale while rejecting the other in its entirety. I agree with Smilansky that the assumption of monism is not tenable, since neither incompatibilism nor compatibilism should be rejected in their entirety. However, I disagree with Smilanksy's particular approach to combining compatibilism and incompatibilism, and after critiquing Smilansky's approach I shall present the case for a fundamental dualism of the kind that I think we should adopt.

The first aspect of Smilansky's fundamental dualism to discuss is his approach to incompatibilism. Where Smilansky is somewhat equivocal I wish to be less so, by affirming that: if the hard determinist contention that it is not possible to be ultimately morally responsible is correct—and the arguments of van Inwagen and Strawson are sufficiently persuasive to conclude that it is—then punishment can never be truly deserved. It is not clear whether Smilansky would assent to the claim in this conditional: on the one hand, he seems to assent to the antecedent, which expresses the conclusion to the hard determinist arguments we encountered in Chapter 4; but on the other hand, he seems to believe that punishment *is* sometimes deserved on compatibilist grounds (for example, the common thief deserves punishment where the kleptomaniac does not). Smilansky's acknowledgement that origination is both a requirement for real moral responsibility and does not exist should force him to affirm that punishment is never truly deserved—yet he confounds our expectations of logical consistency by declaring that whether desert

applies varies according to the situation. To illustrate, Smilansky sketches the case of the lazy waiter, whose inattentiveness towards customers and general lack of effort—despite the fact that he's perfectly capable of acting differently—are punished by the poor tips he receives as a result. In this scenario, Smilansky says, the waiter "does not *deserve* the full tip."[27] We are urged to believe the compatibilist perspective is more salient in this scenario (though it is not clear why), and since the lazy waiter is compatibilistically free to moderate his behaviour in order to make more effort and thereby earn better tips, he deserves to reap the consequences of not doing so.

Quite simply, I see no reason for thinking that the hard determinist perspective does not apply with equal force to this case as indeed it does to all cases. Moreover, once it is applied, the same conclusion will be reached: the agent's behaviour is caused by factors ultimately beyond their control, and hence they cannot ultimately be held morally responsible. So, in the case of the lazy waiter, we can legitimately ask: why is he the sort of person who lacks any inclination to make an effort? Ultimately, we will find that our answer traces back to factors beyond the lazy waiter's control, and for this reason he, like all agents in all scenarios, is neither ultimately responsible for his behaviour, nor deserving of its consequences.

6.4 Consequentialism and Moral Responsibility

After laying out my disagreement here, it may seem as though there is little scope for sympathy with compatibilist distinctions. However, it is less the compatibilist distinctions themselves that I wish to query, but rather their justification. Put simply, whereas Smilansky argues that compatibilist distinctions have crucial *non-consequentialist* ethical significance, I consider that—given that origination is a requirement for moral responsibility—the ethical significance of compatibilist distinctions must perforce be *consequentialist* in nature. We saw in the previous chapter how Vargas seemed to toy with the idea of embracing a consequentialist justification of responsibility norms before eventually rejecting this approach in favour of a (failed) defence of ultimate desert from a compatibilist perspective. I

[27] Smilansky (2002), p. 494.

submit that, once it has been accepted that hard determinist concerns over origination are not misplaced, consequentialism remains the only justification for our responsibility norms on the table. Therefore, not only should we avail ourselves of it, we in fact have no option but to do so if we want to avoid the moral nihilism of abandoning all moral responsibility norms.

Why does a consequentialist understanding and justification of our moral responsibility norms remain the only viable option at this juncture? The reason is that the arguments of van Inwagen and Galen Strawson have found no convincing rebuttal in the mixed views surveyed thus far. In the present case, Smilansky's attempt to persuade us that hard determinism's validity is merely partial and varying is unconvincing, since no real reasons are presented for thinking that its validity may vary according to the case at hand. In fact, in contrast to Smilansky's position, it seems clear that if one accepts that hard determinist arguments invoking the necessity of origination are successful, then one must accept that they are successful in all instances. Rather than offering direct arguments as to why the hard determinist insight is merely partial and varied, Smilansky has sought to appeal to our intuitions through the use of examples such as the case of the lazy waiter mentioned above. These certainly do prompt one into thinking both that compatibilist distinctions are important and that they should be maintained; but they in no way detract from either the force of the hard determinist insight, or from its universally applicable nature.

What is more, our intuitions in response to this and other cases Smilansky discusses can easily be accommodated by a consequentialist approach to moral responsibility norms. Let us take the case of the lazy waiter and see what a consequentialist compatibilism would prescribe here. Since the lazy waiter is clearly capable of doing a better job—capable, that is, in the compatibilist sense of being able to do a better job should he so choose—he is a fitting target of consequentialist punishment, which might very well take the form of receiving a reduction in tips from dissatisfied customers. Despite the fact that he is not ultimately responsible for the lazy behaviour that brings on him the punishment of reduced earnings, the waiter's punishment is nonetheless justified in that it is likely to effect a positive change in

what is reasonably considered to be unacceptable behaviour.[28] Taking the case of the lazy waiter as our exemplum, then, the process of coming to a consequentialist judgment of an agent's moral responsibility seems to involve the asking of the following two questions: (1) Is the agent's behaviour in some way morally unacceptable? (2) Is the agent capable (in the compatibilist sense) of correcting their behaviour? If the answer to both of these questions is 'yes,' then the conditions are met for administering punishment—in the form of censure or concrete sanction—commensurate with the seriousness of the offence.

It should now be clear that the kind of fundamental dualism I am proposing is of quite a different nature to Smilansky's. Whereas Smilansky conceives of a fundamental dualism in which the compatibilist and incompatibilist approaches each have what he calls a "partial and varying validity," and whose applicability varies according to the nature of the case in question, I am arguing that both compatibilism and incompatibilism are wholly and unvaryingly applicable within their own separate spheres.[29] Incompatibilism teaches us that no agent is ever ultimately deserving of the blame or punishment they receive (and, on the flip side, neither does anyone ultimately deserve praise or reward). The rationale for believing this is that true desert requires origination, and origination is a condition that cannot be met—and no amount of compatibilist freedom will ever change that fact. However, while compatibilist freedom does not allow for desert, it can provide support for our moral responsibility practices. We can say that our moral responsibility practices are justified on consequentialist grounds, and that therefore, insofar as compatibilist distinctions have beneficial consequences, we should continue to apply them.

[28] Although I do not here offer an account of how to determine what behaviour counts as unacceptable, note that any such account need *not* necessarily be along consequentialist lines. In other words, I am defending a position of consequentialism solely with regards to the justification of holding people morally responsible. Whether an act is good or bad, moral or immoral, acceptable or unacceptable, is a separate question, and I am happy to let deontologists and consequentialists fight it out between themselves as to whose approach is correct in such an instance. *My* contention is simply that: (1) whether one takes a deontological or consequentialist approach to the issue, the lazy waiter's behaviour here would be widely be regarded as bad, immoral, or unacceptable, and; (2) his behaviour being bad, immoral, or unacceptable is a necessary (though not sufficient) condition to justify his being blamed and/or punished.

[29] Smilansky (2000), p. 286.

6.4.1 Smart's Consequentialism

Consequentialist approaches to justifying moral responsibility norms receive little consideration in the current philosophical literature, and still less support. Perhaps they are perceived as requiring too radical a rethinking of the conditions under which we can hold one another morally responsible—in particular, justifying moral responsibility practices in the absence of even so much as the possibility of desert may be considered problematic. Nonetheless, consequentialist approaches have been championed previously, perhaps most notably by Smart, and it is to his position that we shall now turn in order both to further elucidate the consequentialist approach, and to help answer the charge that moral responsibility requires desert.

In answer to the question of what it is to ascribe responsibility, Smart, like Smilansky, presents us with various examples. One of these examples concerns a schoolboy, Tommy, who has failed to do his homework. There are two possible explanations for this failure, Smart tells us: stupidity or laziness. If the former explanation is correct then there is no place for punishment, while if the latter is correct then punishment may be justified, as Smart explains:

> If Tommy is sufficiently stupid, then it does not matter whether he is exposed to temptation or not exposed to temptation, threatened or not threatened, cajoled or not cajoled. When his negligence is found out, he is not made less likely to repeat it by threats, promises, or punishments. On the other hand, the lazy boy can be influenced in such ways. Whether he does his homework or not is perhaps solely the out-come of environment, but one part of the environment is the threatening schoolmaster.[30]

The point is that "threats and promises, punishment and rewards, the ascription of responsibility and the non-ascription of responsibility" have a clear pragmatic justification regardless of questions of determinism, desert, and origination, and for this reason the lazy Tommy is a fitting target for punishment while the stupid Tommy is not.[31]

Smart argues that the example of the stupid/lazy schoolboy Tommy shows that we must distinguish two ways in which we use the word 'praise': it can either denote the opposite of *blame* (which is what lazy Tommy should receive), or else it can mean the opposite of *dispraise* (which is what stupid Tommy should receive). So, when

[30] Smart (1961), p. 302.
[31] *Ibid.*, p. 302.

used to mean the opposite of dispraise, praise amounts to a positive grading or evaluation, and can apply to moral qualities and actions as well as to physical or mental characteristics such as strength, beauty, intelligence, and so forth (while dispraise, of course, amounts to a *negative* grading or evaluation). Praise of this sort has a primary function and a secondary function, says Smart. Its primary function is to describe what people are like in order to inform others. Speaking perhaps from experience, Smart here offers an example: "To say that one candidate for a lecturership writes clear prose whereas another cannot put a decent sentence together is to help the committee to decide who should be given the lecturership."[32] Naturally enough, then, we come to like being praised and hate to be dispraised, and consequently praise and dispraise come to have an important secondary function: to praise an action is to encourage people to do actions of that class, while dispraise serves to discourage doing such actions.

When the term 'praise' is used to mean the opposite of blame it shares the same meaning as it did when contrasted with dispraise, but with one proviso: to praise a person for an action in this sense is not only to grade it but to imply that the person is responsible for it in the compatibilist sense of being able to do otherwise. Hence, we see that it is appropriate to blame Tommy if his failure to do his homework can be attributed to laziness rather than stupidity. Praising and blaming, urges Smart, should be every bit as dispassionate as praising and dispraising. Although the former involves an ascription of responsibility, this is "perfectly compatible with the recognition that the lazy Tommy is what he is simply as a result of heredity plus environment (and perhaps pure chance)."[33] Therefore, to the extent that people do not praise and blame in a dispassionate and clear-headed way, having as they do the notion that the appropriateness of praise and blame is bound up with their metaphysical (and incoherent) idea of freedom, Smart's proposal represents a revisionist account of moral responsibility.[34]

---

[32] *Ibid.*, p. 304.
[33] Smart (1961), p. 305.
[34] Smart is not proposing a revision to *when* we should hold one another morally responsible (at least, not directly), but rather to the *concept* of moral responsibility. Whereas in most theories of moral responsibility—and indeed in most people's minds—desert is considered a necessary condition for being morally responsible, Smart's consequentialist theory severs this link between desert and moral responsibility.

While there is of course much more to say about what shape a consequentialist justification of our moral responsibility practices should take, Smart's account does give us a rough outline from which to work. More fundamentally, however, I maintain that it is right to think our moral responsibility practices are justified on consequentialist grounds, and that this is therefore how we must respond to Smilansky's call for a fundamental dualism in which both compatibilist and incompatibilist insights are recognised. That this is the right justification can in the first instance be deduced by reflection on the inadmissibility of the alternatives: we are compelled to rule out the possibility of the kind of non-consequentialist justification of compatibilist distinctions that Smilansky favours, since the hard determinist assertion that it is not possible to be ultimately morally responsible has met no successful challenge; equally, we must rule out the possibility of doing as the hard determinist would have us do and rejecting all our moral responsibility practices, since not only would this be likely to have harmful effects for society if it were achievable, but it is in any case beyond our capabilities to do this given our deep-rooted propensity for praising and blaming. With these two options ruled out, consequentialism becomes the only alternative.

Additionally, there is the positive case for consequentialism, which perhaps provides a more satisfying reason for embracing it than does the mere negative appraisal of the alternatives. This positive case will form a large part of the next and final chapter.

6.5 Conclusion

Smilansky's mixed view has provided us with the material to reflect on how we might square the circle of acknowledging the hard determinist's protestations as to the impossibility of ultimate moral responsibility, while still offering a reasonable and reasoned justification of our moral responsibility practices.

Two proposals comprise Smilansky's mixed view: illusionism and fundamental dualism. We saw that illusionism was largely accurate in terms of what it diagnosed—namely, a widespread belief in libertarianism. In this regard, Smilansky's illusionism is in accord with Vargas's revisionism. Additionally,

Smilansky's is right to think that libertarianism is false, and that therefore the majority of people justify their moral responsibility norms by reference to a faulty metaphysical picture. A final point regarding the diagnostic aspect of Smilansky's illusionism is his claim that there is commonly a motivated element to belief in this false picture of libertarianism, the motivation being to evade the troubling reality that the kind of free will the libertarian believes in is a chimera. This is the least secure element of Smilanksy's diagnosis, as an alternative explanation for the widespread adherence to the false belief in libertarianism is available, which is that people are simply mistaken on account of not reflecting sufficiently on the question. As for why people should so commonly be mistaken in the same direction—commonly mistaken, that is, in holding to a belief in folk libertarianism—the reason is our phenomenological experience of agency: in the absence of constraint or coercion, and not having epistemic access to the future, it is natural to think that the 'garden of forking paths' model provides the right metaphysical picture of one's situation. It is only once libertarian agency comes under closer scrutiny that this picture can be seen to be misleading. Nevertheless, Smilansky's suggestion that self-deception is at work cannot be written off, since one can readily perceive a motivation for believing in the faulty metaphysics of libertarianism, which is as follows: if libertarianism really did correctly describe our world then the vexing problem of how to justify our moral responsibility practices would dissolve, because the origination condition would (in certain instances) be met and ultimate responsibility could therefore be ascribed. We thus allowed Smilansky the benefit of the doubt on this matter and declared his diagnosis accurate, self-deception and all, for the sake of argument.

In terms of prescription, however, no such leniency of judgment was extended. Smilansky's argument was that those who are aware of the falsity of libertarian free will must actively seek to obscure this fact from those that are not. The central justification for this "motivated obscurity" regarding the falsity of libertarian free will is that illusion is a condition for supporting the reality of the partial validity of compatibilist distinctions. Without this support from illusion, Smilansky argued, people would be likely to abandon valid compatibilist distinctions and embrace the moral nihilism of hard determinism, heedless of the woeful societal consequences such an embrace would entail. Despite Smilansky's dire warnings concerning the perils of speaking up about the non-existence of libertarian free will, we judged that

his worries were essentially baseless. In fact, it is reasonable to believe instead that our moral responsibility practices are so deeply ingrained that no theoretical consideration would be able to completely dislodge them, no matter how compelling the arguments.[35] In view of this, we should be more concerned about being unable to abandon *unwarranted* habits of praising and blaming than about being unable to maintain *warranted* ones.

The second proposal, Smilansky's fundamental dualism, was judged more successful. In terms of the basic question, at least, we can agree to answering 'yes' to the following: should we abandon the assumption of monism? This, it will be remembered, is the assumption that either compatibilism or incompatibilism is correct in its entirety, and the other is wholly wrong. While we can agree with Smilansky that this is not the case, we must part company with him regarding how to combine compatiblism and incompatibilism, and regarding the underlying justification for this combining. Specifically, we must take issue with two features of Smilansky's fundamental dualism. First, there is Smilansky's failure to acknowledge the implications of the hard determinist's successful argument for the impossibility of desert given the demonstrable impossibility of origination. The main implication is that ultimate responsibility, or true desert, is impossible in *any* situation, and none of Smilansky's attempts to appeal to our compatibilist intuitions are capable of altering that one iota. Second, and following on from the first point, Smilansky presents a *non*-consequentialist justification for the compatibilist distinctions that we are apt to make, such as in the case of the lazy waiter. Naturally, due to the impossibility of true desert, a non-consequentialist justification of these compatibilist distinctions is not available. Instead, we must avail ourselves of a consequentialist justification of these distinctions if we wish to defend them adequately, and Smart's view provides a good starting point for this. In the end, then, we find ourselves with a fundamental dualism in which incompatibilist arguments persuade us that desert is impossible, while compatibilist distinctions provide the basis for a consequentialist justification of our moral responsibility practices. It would be a mistake to dispense

[35] While I believe it is reasonable to believe that no argument would be able to persuade us to abandon our moral responsibility practices, I concede that whether this is the case is an empirical question, and one that could only be definitively answered by running an experiment in which a new incompatibilist education was trialled among a sample of participants.

entirely with either compatibilism or incompatibilism, since the insights of each are applicable in their own sphere.

The form of a workable mixed view of moral responsibility is coming into view now, thanks to our careful scrutiny of the pre-existing mixed views of Vargas, Double, and Smilansky. In the next and final chapter we will continue to present the case for a consequentialist justification of our moral responsibility practices, after which we will go on to consider what has been learned from the mixed views encountered so far. Finally, we will reflect more generally on what we have learned on this journey from metaphysics to morals—from the lofty first principles of the *PSR* to the nitty gritty of how we should assign moral responsibility.

7

A New Mixed View

The final chapter is upon us, and the outline of a plausible theory of free will and moral responsibility is starting to take shape. This theory is one that owes a debt to the insights of both compatibilists and hard determinists, and yet it cannot be said to belong wholly to either camp. It owes to compatibilists the conviction that, under certain conditions, moral responsibility can be justifiably ascribed to individuals, which is to say that we need not attempt to abandon our practices of praising, blaming, punishing and rewarding as the hard determinist wishes to persuade us to do. On the other hand, it owes to hard determinists the acknowledgement that the condition of ultimate desert is one that can never be met, and thus, that while our practices of praising, blaming, punishing and rewarding are sometimes *justified*, this is not to say that they can ever be truly *deserved*. It must be owned that this positing of a distinction between what is justified on the one hand and what is deserved on the other—whereby the former obtains under certain conditions while the latter never obtains—is something that requires further defence and explanation. This will be the first task to embark upon in this chapter, as we consider objections and responses to the kind of consequentialist justification of our moral responsibility norms that I propose we adopt.

Following this, we will reflect on the ways in which the various different elements of the mixed views surveyed are useful to us in constructing this consequentialist justification of moral responsibility. We can embrace some of the insights embodied in the theories of Vargas, Double, and Smilansky without needing to adopt any one of these theories wholesale. The object is to take the best of what we have examined—to affirm what is right about revisionism, fundamental dualism, illusionism, and metaethical subjectivism—and to reject the rest.

Our final task will be to return to where we began, to the *PSR*, in order to reflect on its limits in the light of our discussion of desert, justice, and moral responsibility. Even those of us who embrace the *PSR* to the extent that Spinoza and Leibniz do must concede that it is not applicable in all spheres. It appears that reason has its bounds, after all.

It should be said that the suggestions of this final chapter are speculative and exploratory in nature. Rather than bringing an end to the debate, then, the consequentialist theory outlined is intended merely as a starting point for further discussion.

7.1 Defending a Consequentialist Justification of our Moral Responsibility Norms

The previous chapter offered a brief account of what a consequentialist justification of moral responsibility norms might look like, taking as its inspiration Smart's renowned account. In the first half of this chapter, I wish to build on this account in order to iron out any concerns about whether a coherent case can really be made for such a theory. This will be done by setting out and responding to a couple of commonly-voiced criticisms of consequentialist justifications of our moral responsibility norms, since these threaten to derail the consequentialist project before it can get started. These responses should help to allay worries about the viability of a consequentialist justification, as well as adding flesh to the bones of our mixed view.

7.1.1 Scanlon's Criticism

The first criticism is from Scanlon, who offers a critique of what he calls the Influenceability theory, which is his moniker for Smart's consequentialism. According to Scanlon, the Influenceability theorist's purpose or rationale in administering moral praise and blame is to influence people's behaviour for the better. Consequently, there is "no point in praising or blaming agents who are not (or

were not) susceptible to being influenced by moral suasion," since it is the ability to influence alone that should guide our praising and blaming practices.[1]

After providing an account of the Influenceability theorist's rationale for their moral responsibility ascriptions, Scanlon identifies a potential problem. Some Influenceability theorists might argue that commonsense notions of responsibility should be abandoned to be replaced by their own utilitarian notion—and that is all well and good, supposing one is willing and able to make the difficult case for revising the concept of responsibility in this way. What cannot be argued, however, is that the Influenceability theory provides "a satisfactory account of the notions of moral praiseworthiness and blameworthiness as we now understand them."[2] Scanlon continues: "The usefulness of administering praise or blame depends on too many factors other than the nature of the act in question for there ever to be a good fit between the idea of influenceability and the idea of responsibility which we now employ."[3] So, what on the face of it appears to be a dilemma for the Influenceability theorist—whether to argue that we revise our concept of responsibility so that it accords with their utilitarian outlook, or else to make the case for the Influenceability theory providing a satisfactory account of our current notions of praiseworthiness and blameworthiness—is in fact more of a Hobson's choice situation according to Scanlon, since the latter of these two options is ruled out.

I take no issue with Scanlon's analysis of the problems for the Influenceability theory as described. I agree that, if we were to ascribe praise and blame solely on the basis of whether the target of our ascriptions is susceptible to the influence of moral suasion, then we would have to abandon any hope of maintaining the notion of moral responsibility we currently employ. Further, I do not wish to recommend radical revision to our notion of moral responsibility in order to make it fit with the demands of the Influenceability theory that Scanlon outlines. Instead, I wish to argue for a consequentialist theory that, in contrast to the Influenceability theory, does not require that we act with the conscious aim of maximising utility. In fact, I submit that there is no need to attempt to apply consequentialist thinking at all when making ascriptions of moral responsibility. It is unnecessary—and, indeed, fallacious—to

[1] Scanlon (1995), p. 47.
[2] *Ibid.*, p. 48.
[3] *Ibid.*, p. 48.

conflate the belief that our moral responsibility practices are justified on utilitarian grounds with the belief that we should actively seek to maximise utility when carrying out these moral responsibility practices. The former is what the consequentialist theory that I prescribe asserts, while the latter is no part of it.[4]

It might now be asked: when should we seek to hold people morally responsible, if not when we imagine that doing so will maximise utility? The consequentialist theory I envisage answers that we should hold people morally responsible when and only when two conditions are met: (1) when the agent in question has done something morally wrong, and; (2) when the agent is capable of reforming their behaviour (and hopefully also their attitude) as a result of being held morally responsible.[5] This answer leaves much to be discussed, of course. For each of these conditions we can ask whether the agent, in performing some particular action, meets it (we might have different ideas about what is morally wrong, for example, or we might have different beliefs about the individual's capacity to reflect morally). What can be agreed upon, however, is that these two conditions are commonly met. I contend that when they are met, we are entitled to hold the person in question morally responsible.

The above are the conditions under which I believe we *should* hold one another morally responsible. I am further claiming that we customarily *do* hold people morally responsible under just these conditions. Both of these assertions are of course open to debate, but as a rough outline of how to correctly ascribe moral responsibility it is, I would argue, plausible. Similar criteria for morally responsible agency are presented by various compatibilists such as Fischer and Ravizza, and Haji.[6] What these views have in common with my own is a belief that, for blame and punishment to be justified, it must be the case that some moral transgression has

[4] Arneson also makes this point. In the context of discussing Smart's views on praise and blame, Arneson (2003: p. 240) writes: "For [a consequentialist theory such as Smart's] to work, the agent at the time of praising and blaming probably cannot have in mind the thought that she is behaving strategically to induce good consequences. [T]he conditions that warrant accountability need not be in the mind of someone engaged in an accountability practice." In short, a utilitarian justification of moral responsibility practices need not demand that agents consciously act with utility in mind, nor even suppose that doing so would be possible.

[5] While conditions (1) and (2) can be refined and altered to some extent, they at least give an indication of the kind of theory I wish to defend. Condition (1) indicates that moral responsibility can only be imputed to an agent whose actions violate moral norms, while condition (2) captures the idea that an agent must have sufficient capacity to reflect on their actions if sanctioning or rewarding are to be legitimate.

[6] Fischer and Ravizza (1998); Haji (1998).

taken place, and that the individual committing this transgression has the reflective capacity to respond appropriately to whatever sanction they are given.

My own view is at variance with those of the above-mentioned compatibilists, however, in their respective justifications. Non-consequentialist compatibilists typically argue that agents who meet the above conditions are deserving of blame and punishment, and that this deservingness is what justifies us in meting out our chosen sanction. Against this, I would argue that van Inwagen's and Galen Strawson's arguments force us to conclude that no one ever truly deserves blame, and so clearly the ascription of desert cannot justify blame and punishment. Instead, my position is that we are justified in our practices of blaming and punishing when the two conditions are met, because doing so will ensure the best outcome.[7] To repeat: the claim is emphatically not that we engage in our moral responsibility practices with the explicit intention of achieving the best outcome—our reactive attitudes are far more instinctive and less amenable to shrewd calculation than that. Instead, the claim is that we are liable to blame and punish when the above two conditions are met, and that what justifies us in blaming and punishing under these conditions (whether we know it or not) is the beneficial consequences that are likely to follow from so doing.

To conclude, Scanlon contends that there is a clear divergence between our intuitions regarding when and how to hold each other morally responsible, and what the Influenceability theory recommends. While I accept this contention, and while I furthermore agree that radical revision to our concept of moral responsibility so as to

[7] I acknowledge that the assertion that ascribing moral responsibility in accordance with conditions (1) and (2) will ensure the best outcome is open to debate. Still, while this assertion is unproven, I submit it on the grounds that it is at least plausible (as I will argue), and on the understanding that the conditions are open to revision in the face of evidence that such revision would result in conditions that reap better consequences. In defence of the utility of condition (1), it is plausible to think that failing to prohibit blame and punishment for those who have not done anything morally wrong would have negative consequences, as our sense of justice demands that blame and punishment only be administered in cases of wrongdoing. For instance, although it might be imagined that making an example of the innocent could on occasion be justified on consequentialist grounds—for example, fitting up and hanging an innocent man for some crime as a precautionary warning against others committing the offence for which he is falsely accused—permitting such flagrant injustices would, if discovered, lead to a breakdown of trust in our moral responsibility system and thus do nothing to promote utility. As for condition (2), the case for thinking that this kind of reasons-responsive condition will help promote utility is that only those sufficiently capable of reflecting on their actions are able to learn from being held responsible. As such, if an agent is not capable of reforming their behaviour, we should not expect sanctioning them to produce any benefits—holding someone morally responsible is only of benefit if they are mentally constituted in the right way, and that is what (2) is intended to capture.

make it fit with the Influenceability theory as sketched would be undesirable, I reject Scanlon's implicit assumption that a consequentialist approach to justifying our moral responsibility practices must involve what we might call 'conscious consequentialism': that is, that our moral responsibility ascriptions must be formed while keeping the consequences of these ascriptions in mind. By contrast, the consequentialist theory that I propose does not require us to consider what outcome our praising and blaming will bring. Instead, we should continue to do what I suggest we always have done, which is to apportion praise and blame commensurate with the goodness or badness of the act in question, taking into account the agent's ability to respond appropriately to any sanction. Thankfully, this behaviour comes naturally to us, since, as P.F. Strawson notes, no amount of theorising or metaphysical speculation could conceivably dislodge our reactive attitudes.

7.1.2 Bennett's Criticism

A further criticism of consequentialist justifications of our moral responsibility practices comes from Bennett. His concern is that, while the beneficial consequences that follow from praising and blaming might by happenstance coincide with accountability, consequentialist justifications of moral responsibility norms can never get to the essence of what it is to be accountable. In particular, what consequentialist justifications fail to do justice to, claims Bennett, is the fact that blame-related responses "all involve something like hostility towards the subject."[8] By contrast, proponents of a consequentialist justification of moral responsibility norms can, as Bennett sees it, remain in a "perfectly sunlit frame of mind" while feigning ill feeling for merely therapeutic purposes.[9] While consequentialist theories thus deny that accountability demands reactive feelings, argues Bennett, these theories positively encourage the development of what P.F. Strawson refers to as the 'objective attitude.'[10] Adopting the objective attitude towards oneself and others involves inquiring into how we are structured and how we function – and by means of this inquiry we help to dispel hostile reactive feelings towards the offender as we

[8] Bennett (1980), p. 20.
[9] *Ibid.*, p. 20.
[10] Bennett (1980: p. 21) includes among the reactive attitudes the following: blame, reproach, vilification, resentment, admiration, gratitude, and praise. They are emotional responses, either positive or negative, to a fellow human being, and are bound up with our moral beliefs.

come to view them as simply 'a case.' In failing to recognise that our reactive attitudes are an essential element of accountability, while at the same time encouraging this objectivity of attitude that does much to dispel these reactive attitudes, Bennett concludes that consequentialist justifications of our moral responsibility practices cannot "[do anything] like justice to the real nature of our praise- and blame-related responses."[11]

Bennett's criticism is less easily dealt with than Scanlon's since it contains a kernel of truth, which is that consequentialist justifications of moral responsibility norms fail to do justice to real-life ascriptions of praise and blame. The specific charge is that reactive attitudes such as hostility are essential to any blame-related response, yet proponents of a consequentialist justification of moral responsibility treat them as inessential. The task, therefore, is to defend consequentialist justifications against the accusation that, in treating wrongdoers as cases to be managed, they fail to recognise that true blame requires us to express reactive attitudes.

The first point to make in response to Bennett's criticism must be to concede that a consequentialist justification of our moral responsibility practices cannot capture everything about our ordinary notion of blame. However, Bennett is wrong about precisely what it is that this consequentialist justification fails to capture, since, as I will argue, it is not the case that the consequentialist is unable or unpermitted to view the reactive attitudes as essential to blame. Instead, what the consequentialist theory can rightly be said not to capture about our ordinary notions of blame is the idea that the person on the receiving end deserves their treatment. Still, this need not be fatal to the consequentialist justification since we need not concede that the correct account of what justifies us in our moral responsibility practices must be one that perfectly captures the nature of our current praise- and blame-related responses and beliefs. On the contrary, we can deny that these praise- and blame-related responses and beliefs are in all cases warranted. To the extent that we have a tacit belief that the origination condition is met (and, I would argue, this is the case for the vast majority of us), we will also believe that ultimate desert is possible. This belief is erroneous, so it is of no discredit to the consequentialist justification being proposed that it does not allow for responses that imply that we can be deserving of the praise

[11] *Ibid.*, p. 20.

and blame we receive. Our initial response to Bennett, then, is that he is wrong to assume that the correct justification of our moral responsibility practices must capture all our intuitions about the nature and purpose of accountability. Whereas our current moral responsibility practices are predicated on the notion that it is possible to be deserving of the praise and blame that we receive, the consequentialist justification of moral responsibility norms under discussion here would rightly revise this notion of desert.

Following on from this call for revision to our beliefs about desert, a further possible response to Bennett might be to call for revision to our beliefs about the essentiality of reactive attitudes as well. In other words, we could assent to Bennett's view that reactive attitudes are an essential feature of our current blaming practices, and agree further that no (mere) consequentialist justification of our moral responsibility norms could allow for this fact; yet we could at the same time hold that our current blaming practices should be revised to recognise the fact that reactive attitudes are not essential. However, this is not the line I wish to take, since I think that consequentialism regarding our moral responsibility norms can be defended against Bennett's charge that they treat reactive attitudes as inessential. Recall that, on Bennett's view, consequentialists must express reactive attitudes for merely therapeutic purposes—ill-feeling, hostility, disapprobation and so forth are feigned by the consequentialist in a cynical attempt to discourage wrongdoing, while attitudes such as gratitude and admiration are employed equally cynically in response to virtuous behaviour. While the consequentialist may or may not genuinely feel these reactive attitudes, their only duty is to express them so as to encourage virtue and discourage vice.

We have been here before with Scanlon, as Bennett likewise assumes that those who favour a consequentialist justification of moral responsibility practices must therefore also favour thinking along consequentialist lines when engaging in these practices. The case has already been made that the consequentialist is under no such obligation to perform utilitarian calculations when engaged in moral responsibility practices. The key question is whether the good consequences of our moral responsibility practices *justify* them, not whether good consequences provide the motivation or rationale for acting. The upshot of positing this distinction between the justification of our moral responsibility practices on the one hand and our motivation

or rationale for acting according to moral responsibility norms on the other, is that Bennett's charge that consequentialists will inevitably find themselves obliged to engage in simulated acts of hostility and other such reactive emotions no longer sticks. There need be no calculated element to the consequentialist's ascribing of praise and blame. As has been argued, the criteria for determining whether an agent is a fitting target of (negative) moral responsibility ascriptions are, in the first instance, that they have done something wrong, and, in the second instance, that they are capable of responding appropriately to blame and/or punishment. These criteria, then, both should and typically do provide us with the motivation and rationale to praise, blame, reward, and punish. Further, I submit that we are justified in acting in accordance with them because of the beneficial consequences of doing so. In the light of these facts, Bennett's concern, that those who advocate a consequentialist justification of our moral responsibility norms will be obliged to engage in confected emotional responses for the sake of the beneficial consequences of doing so, no longer looks credible.

The charge that, by being obliged to keep their consequentialist goals in mind, the consequentialist is thus forced to act in emotionally dishonest ways has been considered and rejected, on the grounds that there is simply no need to keep any consequentialist goals in mind when acting. Still, Bennett would probably wish to urge that the revision called for by the consequentialist theory here presented—namely, the revision of our belief in the possibility of desert—renders reactive attitudes inessential anyway. After all, it might be asked, how can we sincerely hold any of the reactive attitudes that are such crucial components of our interpersonal lives if we deny that anyone truly deserves praise or blame?

In response, I think there is a good case for thinking that reactive attitudes can and should retain their essential role in human relations even once a consequentialist justification of our moral responsibility norms and practices has been accepted. One can imagine, for example, feeling disgust at a person whose behaviour is morally objectionable—someone, let us say, who callously refuses to help a vulnerable old age pensioner to cross the road even though this common act of kindness would cost them nothing. It is natural and proper, I would suggest, to believe that this person is not the ultimate source of their own actions and thus not ultimately deserving of blame or punishment, while at the same time feeling—and perhaps also expressing—

disgust at their lack of basic humanity. To take a more positive example, there is no inconsistency of attitude in feeling gratitude towards your friend who has taken considerable time and effort to bake you a delicious birthday cake, while at the same time being aware that they are not ultimately deserving of yours or anyone else's gratitude for their generous character and actions. In fact, it would be absurd to withhold one's gratitude from the friend on the grounds that origination and hence true desert are not possible: the absence of true desert in no way detracts from the fact that her baking you a birthday cake was a very thoughtful and kind act—the very least you could do is to show some gratitude!

The point of these examples is to illustrate the fact that there is no inconsistency in believing that is natural and proper to express the appropriate reactive attitude towards a person who is behaving either virtuously or viciously, while at the same time accepting that ultimate desert (whether in the form of praise and reward, or blame and punishment) is not possible.[12] Nevertheless, while the holding of reactive attitudes should remain an essential part of our interpersonal lives in spite of the denial of the possibility of ultimate desert, it must be accepted that this denial could have fairly radical implications for how we view our moral responsibility practices. While we will still feel disgust and dismay at the criminal's wanton disregard for her victim, I suggest that, in the final analysis, we can no longer view her as truly blameworthy once we become fully aware of the implications of the impossibility of origination. Equally, the altruism of a dear friend, not having its source in her, cannot be considered grounds for deeming her praiseworthy, although we can delight in and feel grateful for her virtues. Their behaviour is, respectively, contemptible and delightful, and it is right that we should censure the former and praise the latter; but neither, ultimately, can be said to deserve either censure or praise.

[12] Pereboom, in response to P.F. Strawson (who was of course responsible for coining the phrase 'reactive attitude'), makes a similar point to the one expressed here, which is that P.F. Strawson is wrong to suppose the hard determinist must forego all reactive attitudes. Where P.F. Strawson argued that the hard determinist is not even entitled to express love on account of the constraints of their theory, Perebom (2001: p. 202) counters: "[M]oral character and action are lovable whether or not they merit praise. Love of another involves, most fundamentally, wishing well for the other, taking on many of the aims and desires of the other as one's own, and a desire to be together with the other. Hard incompatibilism threatens none of this." This is right, I think, and gets to the heart of what is wrong with both P.F. Strawson's and Bennett's charge against hard determinists and those who insist on the impossibility of desert, namely: not all expressions of reactive attitudes depend on a belief in the possibility of agents being deserving of praise and blame.

7.1.3 Questions and Clarifications

With these criticisms answered, it just remains of this section to provide answers to the following three questions, in order to get a better idea of what a plausible theory offering a consequentialist justification of our moral responsibility practices will ask of us: (a) What is it like, psychologically, to live with an understanding of the impossibility of true desert? (b) Is it desirable to attain this kind of understanding? (c) Is it possible to attain it?

Our answer to question (a) will influence our answer to (b), since the desirability of living with an understanding of our lack of deservingness will be influenced by—or perhaps even simply a function of—the experience of having this understanding. A plausible concern is that the knowledge that no-one deserves the blame they receive would make it impossible to persist in the belief that blame and punishment are sometimes justified on consequentialist grounds. Those with such knowledge would therefore feel compelled to relinquish all blaming and punishing behaviours, and this would lead to the destruction of our moral responsibility norms. Although this is a credible line of thought, I cannot agree that an understanding of the absence of true desert would compel a person to abandon all practices of praising, blaming, rewarding, and punishing. As outlined above, I think we can make sense, for example, of blaming and punishing the cruel-hearted man who refuses to help the pensioner across the road, even though we are aware that he himself has done nothing to deserve the cruel character that gives rise to such behaviour. Our moral convictions—that is, our beliefs about what is praiseworthy and blameworthy—should remain intact and untouched by the knowledge of the absence of desert, since I argue that these convictions are guided by conditions (1) and (2) (and knowledge of the absence of desert does not in any way prevent agents continuing to meet these criteria). What I do accept is that the knowledge that true desert is impossible could alter one's understanding of what one is doing when holding another person morally responsible—namely, *not* seeing that person as ultimately deserving of praise, blame, reward, or punishment.

Those who see knowledge of the impossibility of true desert as a threat to our moral responsibility norms will likely answer (b) in the negative: they would argue that it is not desirable to attain this knowledge, on the grounds that the importance of

maintaining our responsibility norms outweigh the benefits of having this knowledge. For those who wish to argue along these lines, a solution might be to embrace illusionism with respect to the reality of the lack of origination and consequent lack of desert. A true understanding of these abstruse and esoteric matters is not worth the price of severe disruption to our system of moral norms, they might say, especially since this system is not in need of repair. Just as Smilansky recommended illusionism with regard to our belief in libertarianism in general, so too is there an argument for maintaining illusion with regard to the issue of origination, which is after all a crucial element within many libertarian theories. All of this is fine as far as it goes. However, if one denies the supposition that the knowledge of the impossibility of true desert will lead to the rejection of our moral responsibility practices, then there is no call for illusionism. My answer to (a) leads me to believe that there is indeed no need for illusionism, and no reason not to think that we should provide an affirmative answer to (b): attaining an understanding of the impossibility of true desert is a desirable state of affairs.

Having provided brief answers to questions (a) and (b), my response to question (c) threatens to render these answers redundant. This is because I doubt that any deep and abiding understanding of the impossibility of true desert is in any case attainable. Taking my cue from P.F. Strawson, I would argue that our habitual ways of thinking about moral responsibility and of responding to one another are so deeply ingrained that they are bound to militate against any profound change in outlook. P.F. Strawson was right to say that, when we ask whether and how we should alter our moral responsibility practices in the light of the truth of determinism, we are in fact imagining that we have "what we cannot have, viz. a choice in this matter."[13] This is not to say that I agree with his further judgement that metaphysical matters (such as the truth of determinism and the possibility or otherwise of origination) are of no importance when it comes to making judgements on ethical issues surrounding free will and moral responsibility—quite the opposite, in fact. P.F. Strawson's insight was one of psychology, not metaphysics, and it leads me to suspect that we must answer question two in the negative: it is unlikely that we could ever really internalise the truth of the impossibility of desert and then respond to others

[13] P.F. Strawson (1982), p. 78.

accordingly.[14] Given my scepticism as to whether it would be possible to live with a deep and abiding understanding of the impossibility of true desert, what it would be like to live this way and whether it would be desirable to do so are questions that consequently assume less importance.

Before we go on to discuss how the various mixed views previously examined relate to the present consequentialist justification of our moral responsibility practices, there is one more issue to consider. This issue is one of terminology, and in particular the clarifying of the use of certain terms within the debate over free will and moral responsibility. A good place to start on this topic is by noting that Galen Strawson's argument, which was presented and appraised in Chapter 4, is in his words an argument for the impossibility of moral responsibility.[15] This argument was judged to be sound; and yet we find ourselves in the present chapter defending a theory that offers a consequentialist justification of our moral responsibility practices. The question this raises is: how can we consistently endorse Galen Strawson's position that moral responsibility is impossible, while we are at the same time advocating a theory that purports to provide a justification for practices that spring from this same notion of moral responsibility?

The key to this conundrum lies in recognising that the notion of moral responsibility to which Galen Strawson appeals, and the one to which a consequentialist justification of our moral responsibility practices appeals, are *not* the same, and that there is therefore no contradiction in denying the possibility of the former while affirming the reality of the latter. Galen Strawson's notion of moral responsibility is one that implies that the agent who is held morally responsible must be deserving of the treatment they receive at the hands of others, whether that be praise, blame, reward, or punishment. The argument for the impossibility of moral responsibility

[14] At least part of my reason for thinking that an intellectually-held belief in the impossibility of origination and desert is unlikely to make much difference to one's everyday judgements of others is that this is what some of the studies examined in Chapter 5 suggest. For example, recall Nichols and Knobe's finding that, while 86% of participants assigned to the abstract condition thought that an agent in a determined universe could never be morally responsible, 72% of those assigned to the concrete condition thought the opposite. Likewise, Roskies and Nichols found that many participants thought that determinism would preclude moral responsibility in some alternate universe, but would not in our own. These studies show that we are much more likely to have compatibilist intuitions when presented with concrete, this-world scenarios, and they also suggest that real-world judgements of moral responsibility are liable to persist regardless of what metaphysical views we hold of our world.

[15] That this is so is nowhere clearer than in Galen Strawson's (1994) paper, entitled: "The Impossibility of Moral Responsibility."

states (quite correctly) that the impossibility of origination entails that no-one can truly deserve either praise and reward or blame and punishment, and, given that the present notion of moral responsibility requires desert, we can infer that it is impossible for anyone ever to be morally responsible. The notion that a legitimate judgement of moral responsibility requires that the agent deserves whatever treatment such a judgement entails seems to me to be in line with what Vargas would call our 'folk concept' of moral responsibility. In other words, desert is commonly regarded as a precondition for moral responsibility, such that, in the absence of desert, it would no longer be considered possible to legitimately ascribe moral responsibility.

The kind of consequentialist justification of our moral responsibility practices under discussion here, however, does not require desert. On the contrary, the present consequentialist justification acknowledges the impossibility of desert, and makes the case that we have grounds for maintaining our moral responsibility practices regardless. Of course, making this case involves defending a fairly fundamental revision to our notion of moral responsibility, a revision that calls for the severance of the link between moral responsibility ascriptions and desert. Still, my hope is that what has been written about consequentialist justifications of moral responsibility norms thus far demonstrates the possibility of providing not just a plausible account of how we should think about moral responsibility, but one that does not depend on a condition (namely, true desert) that cannot be met.

To conclude, unspecified terms are a potential source of confusion when discussing issues of free will and moral responsibility. In this instance, disambiguation between two different usages of the term 'morally responsible' has helped to clarify how ostensibly conflicting beliefs can in fact be simultaneously affirmed. On the one hand, I agree with Galen Strawson that it is not possible for any agent to be morally responsible, so long as moral responsibility is understood to require deservingness (which most people, I would argue, believe it does require). On the other hand, I propose that we operate according to a notion of moral responsibility that does not see desert as a precondition for legitimate moral responsibility ascriptions. The consequentialist justification that has been sketched so far demonstrates that the rending of desert and moral responsibility is possible: given the impossibility of true desert and the importance of our moral responsibility practices to our daily lives, it is

also necessary. In the following section, we will further discuss what kind of revision to our concept of moral responsibility (and perhaps also of our practices) a consequentialist justification calls for, and we will also consider how other aspects of mixed views—illusionism, fundamental dualism, metaethical subjectivism—are of use in establishing a consequentialist justification of moral responsibility norms.

7.2 The Debt to Other Mixed Views

We have already touched on how certain aspects of mixed views relate to the consequentialist justification of moral responsibility norms being propounded here. In this section, we will go through the various features of the mixed views surveyed in the previous two chapters to see what use they are to us in shaping our consequentialist theory.

7.2.1 Revisionism

As I have argued already, it is generally accepted that we can be ultimately deserving of praise and blame when in fact the arguments of van Inwagen and Galen Strawson should force us to conclude that this is mere illusion. In consequence, any clear understanding of what it is to justly hold a person morally responsible must involve the recognition that ultimate desert is impossible, and that moral responsibility must instead be justified on consequentialist grounds. Embracing this consequentialist justification of our ascriptions of moral responsibility therefore involves revision to our concept of moral responsibility: we deceive ourselves if we see ascriptions of moral responsibility as implying desert, because no-one ever ultimately deserves any of the praise, blame, reward, or punishment that they are liable to receive. The concepts of moral responsibility and of desert are thus no longer wedded to one another: moral responsibility ascriptions are seen to derive their justification from their beneficial consequences, while any ascription of desert must simply be mistaken.

As we have already seen, revising our understanding of the concept of moral responsibility in this way is potentially confusing. On the one hand, we must deny

the application of our ordinary, 'folk' notion of moral responsibility, yet on the other hand, we must affirm the reality of a revised notion of moral responsibility that does not require desert. So, whether we should affirm or deny the possibility of moral responsibility really depends on whether the term is understood in its ordinary, 'folk' sense, or else in the revised sense being suggested here. Unless it is clear in what sense the term is being used, confusion is apt to arise.

What are the similarities and differences between Vargas's revisionism and my own? The most important features our theories have in common is that we both agree that the current folk concept of free will (of the kind required for moral responsibility) is libertarian, and we both seek to revise this concept by suggesting compatibilist conditions under which our moral responsibility practices are justified. Where they differ quite significantly, however, is on the question of desert: whereas Vargas believes that praise and blame are not only *justified* under the right compatibilist conditions, but that praise and blame can also be truly *deserved* under these same conditions. My belief in the soundness of Galen Strawson's and van Inwagen's arguments prevents me from assenting with Vargas on this question of desert, and I differ from him further in my belief that consequentialism must be embraced in order to justify our moral responsibility practices. In a sense, then, my own revisionism is more radical than that proposed by Vargas: where we both agree that libertarian folk concept of free will should be revised in favour of a compatibilist understanding, I maintain further that our concept of moral responsibility needs revision to reflect the fact that true desert is impossible, and that our system of moral responsibility norms thus requires a consequentialist justification.

While the adoption of a consequentialist justification of our moral responsibility norms requires us to change our *concept* of moral responsibility, it is less clear what impact, if any, its adoption would have on our moral responsibility *practices*: conditions (1) and (2) are essentially compatibilist, and it is a reasonable assertion that conditions much like these govern our current moral responsibility practices; on the other hand - and in opposition to standard compatibilist theories—these two conditions only cover when ascriptions of blame are *justified* and not when ascriptions of blame are *deserved*, and this absence of desert might still be thought to affect practices. For instance, there is a case for thinking that an awareness of the impossibility of desert would alter the way in which one ascribes moral

responsibility to others: for instance, perhaps a deep understanding of the contingency of the causes of the murderer's vicious character would reduce one's inclination to hold him morally responsible. Still, the thought that this revision of our concept of moral responsibility would have any significant effect on our moral responsibility practices depends on the assumption that we are actually capable of internalising the fact that desert is impossible and then acting accordingly. Even if one accepts the arguments for consequentialism, I have expressed my doubts about whether any intellectually-held belief in the impossibility of origination and desert could ever translate into a deep and abiding understanding that would guide one's behaviour towards others.[16] As such, my suspicion is that even if the revised concept of moral responsibility proposed here were widely accepted, it would be unlikely to have any great impact on our moral responsibility practices.[17]

These comments on the effects of adopting a consequentialist justification of our moral responsibility norms are unavoidably speculative, of course, since the effects of the theory could only be known if it were widely embraced. It is possible, therefore, that embracing a consequentialist justification would profoundly affect our moral responsibility practices, in which case we should ask whether the practical consequences of adopting such a justification are desirable. If the answer to this is 'no,' then there might be a case for appropriating elements of Smilansky's illusionism. This is what we shall consider next.

7.2.2 Illusionism

I argued in the previous chapter that Smilansky's illusionism is unnecessary, largely on the grounds that widespread knowledge of the falsity of libertarian free will is unlikely to have the serious negative repercussions that Smilansky envisages.

[16] These doubts stem from consideration of the findings of various studies in Chapter 5 (e.g. by Nichols and Knobe, Roskies and Nichols); P.F. Strawson's persuasive case for thinking that metaphysical beliefs are incapable of altering our moral responsibility practices; and my own first-hand experience of continuing to praise and blame both myself and others despite my conviction that van Inwagen's and Galen Strawson's arguments are successful.

[17] Even if it were to transpire that we are capable of fully internalising the fact that ultimate desert is impossible, I think we would still be inclined to retain our current moral responsibility practices in much the same form that they are now. It seems to me that our current moral responsibility practices are largely effective in encouraging virtuous behaviour, and so, even if we no longer believe in the reality of desert, there is still good reason to persist with these practices.

However, taking Smilansky's illusionism as inspiration, it is possible to make a case for maintaining a motivated obscurity with regards to the impossibility of ultimate desert. While Smilansky advocates illusionism with respect to libertarian free will in order to stave off a perceived threat of moral nihilism, the argument in the case of illusionism with respect to the lack of ultimate desert is that this would help ensure that our moral responsibility practices are effectual in achieving their consequentialist purpose. So, if knowledge of the impossibility of ultimate desert should prevent people from engaging in moral responsibility practices that have beneficial consequences, this might provide grounds for promoting illusionism with respect to the impossibility of ultimate desert.

There is some tension between the revisionism sketched above and the illusionism described here: the revisionist urges us to obtain a greater degree of insight into the concept of moral responsibility by revising this concept in such a way that we expunge the false belief that ultimate desert is a reality; the illusionist, meanwhile, points to the dangers of gaining greater insight into the concept of moral responsibility, and seeks instead to persuade us that our mistaken belief in ultimate desert is in fact an untruth whose convenience warrants its retention.

The motivation for the kind of illusionism under discussion is that, without the background belief in ultimate desert, attempts to hold people morally responsible will be deprived of the force they need for ensuring effective outcomes. There are two sides to this: on the side of what we might call the 'blamer,' the allegation is that their knowledge that the person being blamed cannot truly deserve their treatment is bound to mitigate the force of the blame being administered; while on the side of the 'blamee,' it is claimed that their knowledge of the same will plausibly make them less inclined towards any attempt to mend their ways.[18] Illusionism with regard to belief in ultimate desert might therefore be seen as a useful fiction, without which we would struggle to uphold our moral responsibility practices.

There are two basic points to make in response to this call for illusionism with regard to the impossibility of ultimate desert, both of which have been touched on already but which are nonetheless worth restating. The first is that the kind of motivated self-

[18] Although I use the labels 'blamer' and 'blamee' in making the point, note that the same thought applies equally to cases of praising, rewarding, and punishing.

deception being promulgated here is unlikely to be necessary, on account of the fact that there is little chance of us coming to any lasting appreciation of our mutual lack of ultimate deservingness for the good and bad things we have done—in short, the illusion is already so powerful that it does not need bolstering with motivated self-deception. Since our belief in ultimate desert is so deep-rooted that we will likely continue to view ourselves and others as though this illusion is real, there is simply no call for concern over what the consequences for our moral responsibility practices might be if we were stripped of this illusion.[19]

The second point is that, if it is in fact possible to thoroughly purge ourselves of the illusion of ultimate desert, there is a good case for thinking that doing so would be desirable. One would doubtless expect the nature of praising and blaming to change as we, as members of society, come to internalise the realisation that ultimate desert does not exist. This change is to be desired, however, in that we will view each other with a clearer eye, untainted by false notions of desert. Crucially, the absence of ultimate desert need not in any way diminish our distaste for immoral and reprehensible behaviour, which is perhaps not sufficiently appreciated by those with misgivings about consequentialist justifications of moral responsibility norms. Equally crucially, there is no reason to think that the targets of our moral responsibility ascriptions would change radically if the illusion of desert were eradicated. On the contrary, insofar as our present moral responsibility practices have beneficial consequences—and, in my estimation, they almost invariably do—the realisation of the impossibility of desert would have no significant effect on *when* we hold one another morally responsible. Rather, it is the *quality* of our praise and blame that we could expect to see altered, as we come to recognise that Dennett's rhetorical question: "who more deserves to be despised than someone utterly despicable?" merits a more nuanced response than he himself imagines.[20]

[19] See fn. 14 and fn. 16 for reasons why I doubt we are capable of transcending the illusion of ultimate desert.
[20] Dennett (1984), p. 167.

7.2.3 Fundamental Dualism

Smilansky is right to believe that compatibilists and hard determinists each have important insights into the free will debate. However, as I argued in Chapter 6, we must reject Smilansky's own understanding of what a fundamental dualism should look like, as his account conflicts with the kind of consequentialist justification of moral responsibility norms being sketched here. In particular, what cannot be accepted is Smilansky's claim that "very often both [compatibilist and hard determinist] perspectives are important simultaneously and imply contrary judgements," the implication being that, when making judgements of moral responsibility, we need to be balancing compatibilist distinctions for grounding moral responsibility and hard determinist avowals that no such distinctions exist.[21] In certain cases, Smilansky claims, compatibilist considerations will seem more pertinent (for example, in the case of the lazy waiter, discussed in the previous chapter), while in other cases, hard determinist ones will come to the fore—and it is a matter of judgement to strike the right balance between the two.

While this account of what fundamental dualism should look like views compatibilism and hard determinism as each being partially adequate within the same sphere, the fundamental dualism that I propose we embrace sees each as wholly adequate and applicable within their own, separate spheres. When it comes to making judgements of moral responsibility, it is compatibilist distinctions alone that provide our guide, and it is the beneficial consequences of making these distinctions that justifies us in making them. This approach contrasts with Smilansky's on both of these counts: on the first count, this approach advises that there is no need to balance the conflicting views of compatibilists and hard determinists when seeking to ascribe moral responsibility—instead, we work solely on the basis of compatibilist conditions (1) and (2); on the second count, Smilansky's assertion that compatibilist distinctions have "crucial (nonconsequentialist) ethical significance" is denied, since, given the acknowledged success of hard determinist arguments against the possibility of desert, the only grounds on which these distinctions can be justified are consequentialist ones.[22]

---

[21] Smilansky (2000), p. 286.
[22] Smilansky (2002), p. 493.

Compatibilist distinctions are thus seen to provide our sole guide in the sphere of making moral responsibility judgements, meaning there is no need to weigh the opposing perspective of hard determinism in the balance each time we ascribe moral responsibility to one another. However, the hard determinist perspective is applicable within its own sphere, which is our understanding of desert. What a proper recognition of the hard determinist arguments grants us is the realisation that ultimate desert is impossible: that is to say, that all blame and punishment – and indeed all praise and reward—is undeserved.[23] If we are willing to accept the intuition that origination is a requirement for desert (as I believe we should), then this is the inescapable conclusion of van Inwagen's and Galen Strawson's arguments—and so we see that desert is not a requirement for moral responsibility practices that are guided by compatibilist distinctions, and justified by their beneficial consequences.

A couple of advantages to this reconstitution of Smilansky's fundamental dualism suggest themselves. For one thing, the fundamental dualism being proposed here does away with the need to deny the universal applicability of the hard determinist insight, and instead embraces the reality that no agent ever deserves the praise, blame, reward, or punishment they receive. We saw how Smilansky sought unsuccessfully to deny this point by means of his sketch of the lazy waiter. Unconvincing denials of this sort are not necessary with our present account of fundamental dualism. A second point is that our reworked fundamental dualism avoids us having to embrace what could either be referred to politely as paradox, or else, rather less diplomatically, dismissed as contradiction. Smilansky sees compatibilism and hard determinism as applicable within the same sphere, with the ultimate perspective of hard determinism always threatening to render compatibilist distinctions irrelevant. By placing hard determinism and compatibilism within their own, separate spheres, we are thus no longer obliged to make attempts to reconcile these two irreconcilable positions.

[23] In being ultimately undeserved, it is also in a sense unfair that some are praised and rewarded while others are blamed and punished—the lottery of life determines that some enjoy the former while others must suffer the latter, and that this is so is not fair. However, praise and blame and their correlates are justified on account of their beneficial consequences, and so it is fair to praise and blame when this justification warrants it. Perhaps we can say that life itself is unfair, and our system of moral responsibility, through which we strive to achieve the best consequences, represents the least bad of various bad options.

7.2.4 Metaethical Subjectivism

Metaethical subjectivism, as propounded by Double, was judged to be wrong. In particular, Double is wrong to suggest that there are no objective truths concerning free will and moral responsibility.

Nevertheless, it was found that Double's 'prudential justification' for metaethical subjectivism does offer some interesting lessons for those of us engaged in theorising about moral responsibility. This prudential justification, it will be recalled, stated that it would be preferable for us to hedge our bets by *not* adopting any particular lower-level theory on the grounds the risk of picking the wrong one is too great. Double argues that by retaining and acting on our current, conflicting moral intuitions about free will and moral responsibility we thereby guarantee our intuitions are wrong only some of the time, whereas if we adopt some particular lower-level theory we run the risk of being wrong all the time.[24]

Double's insight is that there are risks in attempting to make our intuitions conform to theory. Although he overstates his case significantly by implying that choosing which lower-level theory to endorse is a matter of mere guesswork, it is true that we cannot be certain that our choice of lower-level theory is correct. It is an advantage of the consequentialist justification of moral responsibility being defended, then, that it attempts to make sense of both hard determinist and compatibilist intuitions: it provides a framework for understanding hard determinist intuitions about the impossibility of true desert, as well as compatibilist intuitions about the necessity of holding each other morally responsible under certain conditions. Further, this consequentialist justification does not call for radical changes in our moral responsibility practices, even if the theory itself is quite radical.

In summary, Double's prudential justification for his metaethical subjectivism is of value not in persuading us of the necessity of adopting his theory, but rather in reminding us to consider the risks of abandoning our pre-theoretical intuitions for the sake of some particular lower-level theory. If, as I hope, the consequentialist justification presented is correct then this reminder is simply unnecessary; and if it is

[24] Double's strategy does not diminish the *average* likelihood of our intuitions being wrong, of course—it merely averts the danger of them being utterly wayward on account of having chosen the wrong lower-level theory.

incorrect, then at least it should preserve our intuitions well enough to allay concerns that its implementation might lead to injustices.

7.3 The *PSR* and Its Limits

We have now sketched the outlines of a viable mixed view of moral responsibility, one that accepts hard determinist arguments against the possibility of desert, but at the same time makes a case for maintaining our moral responsibility norms by drawing compatibilist distinctions for consequentialist reasons. There is a lot that still remains uncertain with this mixed view—for example, it is not clear whether it is possible to truly internalise our own and others' lack of desert, nor whether doing so would adversely affect our ability to maintain our responsibility norms—but we must be content to hold off for the time being on trying to provide definitive answers to these issues. In any case, enough flesh has been put on the bones to show that there is an answer to the problem of whether we have the free will required for moral responsibility that relies neither on a perfunctory and shallow dismissal of the hard determinists' arguments, nor on an implausible and unappealing call for the wholesale rejection of our moral responsibility practices.

In this final section, then, let us return to where we began, with the *PSR*, in order to reflect upon what the present theory has to tell us about the principle's limits. It was argued in Chapter 2 that we should adopt a presumption in its favour, meaning that, in the absence of proof to the contrary (in the form of either incontrovertible evidence for some event or occurrence that lacked a sufficient reason, or else an argument for the falsity of the principle whose soundness was beyond doubt), we should assume that the *PSR* is universally applicable. Since no proof of the falsity of the *PSR* was found, the presumption in its favour remained secure, and this in turn gave us reason to embrace determinism, whose truth it was thought to entail. The conclusion of our investigation was thus: there is a reason for everything that happens in terms of prior causes that determine its happening.

However, as all-embracing as this statement of the *PSR*'s applicability may appear, one can in fact imagine it having a still wider scope than this. For instance, it might also supply *moral* reasons: for the facts of our character, for instance, and for the

circumstances into which we were born. We can wonder at the reason that some people are born with all the advantages that good genes and environment bring, while others are condemned to suffer from the lack of these advantages; but it seems that such a question, although perfectly intelligible, has no answer.[25] On the question of moral reasons for facts about our world, then, it seems that we come up against the limits of the *PSR*: there simply are no moral reasons that explain the vast discrepancies in quality of life within and across communities and cultures—and hence neither is there is any justice to these discrepancies.[26]

This absence of moral reasons for facts about our world gets to the heart of what is wrong with our tacit belief that true desert really is possible. In accepting the reality of true desert as most of us instinctively do, we erroneously ascribe responsibility to ourselves and each other for what is in fact the unchosen and adventitious raw material of our character and environment. We think and behave as though the praise, blame, reward, and punishment that we and others receive on account of our moral strengths and failings—and which are, after all, the product of the contingencies of our circumstances and character—are earned; but these could only be earned if there were some satisfactory—that is to say, moral—explanation for the contingencies of circumstance and character. There is no explanation for these, however, and so this absence of a sufficient reason for the moral facts obliges us to reconsider our instinctive belief in desert.

---

[25] Some have certainly attempted to provide a satisfying answer to this question. Hindus and Buddhists, for example, would certainly argue that there are moral reasons that explain your circumstances in this life. The doctrine of karma, common to both of these religions, holds that we reap what we sow across countless lifetimes of rebirths, with the good or bad karma accrued in one life accounting for the blessings and hardships of the next. While I do not dismiss out of hand the possibility that the metaphysical doctrine of karma is correct, I do dismiss the idea that embracing this doctrine would, as a consequence, permit one to believe that praise, blame, reward and punishment can be truly deserved. If one's character and circumstances in this lifetime are supposed to be one's just deserts for the actions of the previous lifetime, the problem of explaining desert is merely deferred, since one can always ask: what did I do to deserve my character and circumstances from that previous lifetime? An infinite explanatory chain is set in process, whereby each incarnation is said to be deserved on the basis of the previous one. The condition of origination—the condition, that is, of being the source of one's actions—can no more be met across countless lifetimes than it could within the one lifetime, and for this reason karma and rebirth are incapable of offering a means of moral explication for the contrasting fortunes of the mass of humanity.

[26] Perhaps there are other meaningful questions that arguably have no sufficient reason, such as: why is there something rather than nothing? Why do I exist rather than not exist? On the question of why there is something rather than nothing, recall from the first chapter how Russell (in Hick (1964: p. 175)) argues that we need not suppose that there is any explanation for the universe as a whole. Rather, the universe is "just there, and that's all."

It is this limit to the applicability of the *PSR* that forces us to confront a fact that we would rather avoid: that the world in which we live is inherently unjust. If asked, most (barring the Hindus and Buddhists among us) would accept that it is wrong-headed to demand a moral reason as to why some people are born with good characters into an environment in which they can flourish, while others are born with bad characters and in bad circumstances. In other words, we know that the raw material of our personality and the environment in which we are born is purely a matter of our good or bad luck. Still, this knowledge rarely translates into the kind of deep understanding that is required in our everyday relations, an understanding that does not view people as deserving of the blame and punishment they must face as a consequence of their wrongdoing. It is perhaps simply easier to ignore this lack of ultimate desert—to disregard the inherent injustice that is the accident of birth, and instead to behave as though the virtuous and the vicious alike have earned their fate. Perhaps there is also a tacit belief among some that the moral responsibility system is contingent upon us living in a just world, and so any acknowledgement that one's birth is a lottery is ignored whenever possible in order to protect and maintain this system.

However, it is a mistake to think that our moral responsibility system requires that we live in a just world, or even that the system itself must be perfectly just. Instead, we must accept that it is an unjust tool for use in an unjust world. In stating that our moral responsibility system is an unjust tool, I do not wish to suggest that it either should or need be in any way arbitrary or inconsistent—it is merely to point out the injustice that blame and punishment will surely follow for those born with poor character and in unfortunate circumstances. As explained already, punishment and reward can be justified for the positive changes they are able to bring about; but it is ultimately unjust that one person is in a position to be rewarded while another person must face punishment.

To say that the *PSR* has its limits, and that the apparent accident of our births provides evidence of this, is essentially to say that there are some questions that make sense and yet have no answer. However, perhaps this is the wrong way of looking at things. Instead, we can say that certain moral questions—for example, "why do bad things happen to good people?" and, "why are some people born with all the advantages life can offer while others must overcome the most formidable

obstacles?"—are unanswerable simply because they rely on a false supposition. In the case of these two questions, the supposition is that our world must be inherently just—that is, both questions demand reasons that will resolve what appear to be injustices, and provide a moral explanation for suffering and misfortune. Once the falsity of this supposition has been accepted, we can say that there *is* no reason that bad things happen to good people, or that some have all the advantages while others have only misfortune. Better still, we can say that the reason for these things is that we are born into a world that is indifferent to our sense of justice, and so it is a mistake to look for reasons to explain away injustice. Seen in this way there is actually no question to answer, and so it is a mistake to view it as a limit to the *PSR* that it is unable to provide one.

In summation, while the *PSR* certainly applies to our universe and its physical order (in that it tells us that everything within our universe comes to exist through necessitating causes), it manifestly does not extend to the moral order in the way we might wish. That is to say, the *PSR* does not explain how we could deserve our happiness and sorrows, for the simple reason that these *are* not deserved. Nonetheless, in the absence of reasons for our respective fates in life, we must impose what moral order we can in order to make the best of our world. This is something we already do, of course, although in our confusion we are liable to assume that the targets of our moral responsibility judgements deserve these judgements in a way they could only if our universe were to have the kind of moral order it so evidently lacks. It is clear that the *PSR* is not applicable to the moral order in the same way that it is to the physical order—the apportioning of good and bad fortune is indeed a mere matter of fortune. What we must recognise, therefore, is that, to the extent that our moral responsibility judgments are justified, they are justified purely on utilitarian grounds. By balancing our clear-eyed, compatibilist, utilitarian judgments with the equally vivid apprehension of the impossibility of ultimate moral responsibility, there is hope that we might develop an interpersonal life that upholds moral standards without lacking compassion.

Conclusion

The arguments of the preceding chapters have yielded the following four principal conclusions: that the *PSR* is true; that the *PSR* entails the truth of determinism; that, whether or not one accepts the truth of determinism, ultimate moral responsibility is not possible; and that the best (and perhaps only) way of justifying our moral responsibility practices is to adopt a consequentialist compatibilist theory. Additionally, empirical evidence has been collated which suggests that most of us are inclined towards adopting a folk libertarian position with regards to free will and moral responsibility—in other words, most of us believe this world to be both indeterministic and one in which we are fully morally responsible for at least some of our actions. This evidence provides some support to the mixed views of Vargas and Smilansky, both of which rely on the assumption that folk libertarianism is pervasive. However, it also reveals that our folk beliefs are perhaps more nuanced than these mixed views allow: rather than attempting to ascertain simply whether the majority of people are compatibilists or incompatibilists, majority opinion is better understood as conforming to a hierarchy of beliefs in which belief in moral responsibility is non-negotiable, while belief in libertarianism is strongly-held although not intractable.

Various objections towards both the methods and conclusions of this thesis are to be anticipated, and I will here outline what I consider the three most significant among these. The first objection concerns the contention that the *PSR* is true. Where I saw arguments on both sides that were inconclusive but that gave us reason to adopt a presumption in favour of the *PSR*, others would doubtless adopt a presumption in favour of its falsity, or perhaps even find the arguments of Hume or van Inwagen decisive in proving its falsity. All I can say to such people is to acknowledge that I do not expect the arguments in the *PSR*'s favour to appear as compelling to others as they do to me. Schopenhauer's admonition that it is foolish to seek to provide proof for a principle that is itself in some sense fundamental was never far from my consciousness as I sought to do just that, and my suspicion remains that neither

arguments for nor against would be likely to change anyone's pre-existing intuitions on the *PSR*. My feeling, then, is that if one finds the arguments for the *PSR* compelling then that is probably only because one was convinced of the truth of the principle itself beforehand. Likewise, arguments against the *PSR* will only appear compelling to those who have a predisposition towards rejecting the principle.

Despite the acknowledged difficulty of persuading anyone to accept the *PSR* if they are not already inclined to do so, I take some solace from the fact that others have attempted, however unprofitably, to do the same: Della Rocca, for instance, concedes that his own attempt to state the case for the *PSR* must look "quixotic" given the principle's current unpopularity.[1] For all that my own attempt to seek for a proof of the *PSR* may appear equally foolhardy, I hope that the discussion of the first two chapters at least leads others to reflect on the extent to which explanations for things and events are required and why.

A second objection to consider is one that is likely to be raised by all those who are non-consequentialist compatibilists—a sizable number of people, in other words—and it is that I am mistaken in considering it necessary to meet the origination condition if we are to ascribe true or ultimate moral responsibility to one another. Chapter 4 examined the arguments of van Inwagen and Galen Strawson and concluded that both were successful; but the success of these arguments is predicated on the belief that justification for our moral responsibility practices requires that the origination condition be met. In the case of van Inwagen's argument, the key premise that expresses this commitment to the origination condition states: "No one has power over the fact that the facts of the past and the laws of nature entail every fact of the future, including one's own actions."[2] In the case of Galen Strawson's argument, the key premise reads: "If you do what you do because of the way you are, then in order to be ultimately morally responsible for what you do you must be ultimately responsible for the way you are."[3] Each in their own way expresses the idea that, if an action originates from some external source over which the agent has no control, then that is not an action for which the agent can be responsible.

[1] Della Rocca (2010), p. 1.
[2] Van Inwagen (1983), p. 222.
[3] Strawson (2002), p. 443.

Non-consequentialist compatibilists must—and are happy to—deny the need to meet the origination condition, and indeed the case against it was at the heart of many of the compatiblist responses considered in Chapter 4. Slote, for example, claimed that, whether an action originates in a causal chain that begins outside of the agent is of little importance so long as the desires, abilities, and so forth of that agent are engaged in the right ways. The agent's action cannot be said to be unavoidable and inevitable despite the lack of origination, and hence they retain moral responsibility for it. Similarly, Hurley contends that "a person need not be responsible for being what he is in order to be responsible for choices that are determined by what he is."[4] Even though we are not responsible for our characters, Hurley claims that this is not to say that we cannot be responsible for the actions that spring from our characters. Once more, this amounts to a repudiation of the need for origination for justifying our moral responsibility ascriptions.

I am not convinced that much more can be done besides noting this divergence in intuitions regarding the importance or otherwise of the origination condition. As far as I am concerned, there is a clear case for siding with van Inwagen in thinking that one cannot be responsible for something that is the unavoidable consequence of facts about the past before one's birth plus the laws of nature. Equally, I am entirely sympathetic to Galen Strawson's claim that, if you do what you do because of the way you are, then in order to be ultimately morally responsible for what you do you must be ultimately responsible for the way you are.

We seem to have hit philosophical bedrock at this point, where all that can be said is that some share these intuitions while others—Slote and Hurley among them—disagree. Certainly, I can see no further arguments from either Slote or Hurley to back up their assertions that the origination condition need not be met, and neither can I think of how to provide further arguments in favour of its necessity. Either one accepts that it is necessary—which, I contend, it manifestly is—or else one rejects this. Perhaps what both this and the previous objection illustrate, then, is the extent to which intuitions underpin so much of our philosophical thinking. While philosophy is ostensibly all about arguments—and these are of course extremely important—the fact is that, as van Inwagen notes with reference to his intuition in

[4] Hurley (2000), p. 30.

favour of Davidson's criterion of individuation, "arguments must come to an end somewhere."[5] When they do, we are often left with intuitions for which we find no further justification beyond simply being able to cite their intuitive appeal. Where Davidson's criterion of individuation is for van Inwagen one such intuition without justification, so is the origination condition for me. And, in the light of Schopenhauer's admonition discussed above, perhaps the *PSR* should be considered to be another.

A final objection to anticipate concerns the case made for a consequentialist compatibilism, which is introduced in Chapter 6 and developed further in Chapter 7. In particular, the case for consequentialism is far from being fully worked-out in this thesis, and it might be felt that there remains a certain tension and lack of clarity in the arguments. This is perhaps most apparent in the account provided in Chapter 7 of how adopting a consequentialist justification of our moral responsibility practices should affect how we respond to and view one another. I made the case that, notwithstanding Bennett's arguments to the contrary, we can continue to hold reactive attitudes towards each other even once we have embraced the notion that praise, blame, punishment and reward can never be truly deserved. So, we can still feel gratitude for our friend who goes to the trouble of baking us a birthday cake despite our knowing that gratitude can never be ultimately deserved, and we can still punish the thief for his thefts even though we understand that no punishment can ever be ultimately deserved either. The objection to this is that there is an insuperable tension between our knowledge that moral responsibility ascriptions can never be deserved, and the claim that we can persist in holding reactive attitudes in a way that seems to imply that they are deserved.

I confess that the issue of what reactive attitudes are rational and/or permissible is a thorny one. Still, I believe it is possible to provide a coherent picture of what embracing a consequentialist justification of moral responsibility norms would entail for our interpersonal lives. The first thing to note is that exactly what it entails depends on the extent to which we are able to behave as fully rational beings—beings that, *inter alia*, are capable of acting with a thorough understanding of the impossibility of desert. I have already expressed doubts about our ability to do this,

[5] Van Inwagen (1983), p. 169.

and as such I think it likely that we will always have a residual tendency to treat one another as though we truly deserve praise, blame, reward, and punishment. Given this assumption, the cake baking and thieving examples are potentially misleading if taken to imply that the consequentialist theory presented assumes that we will be able to form a complete understanding of the impossibility of desert.

Imagining, however, that we were able to act as fully rational beings: how then would we react to our cake-baking friend and the thief? I think the answer must be that we could only feign praise and blame, and that we would feign these attitudes for consequentialist reasons—that is, in order to encourage acts of cake-baking, and discourage acts of thievery. Of course, a fully rational agent need not feign attitudes such as delight and disgust since these attitudes do not imply any belief in the moral culpability of the agent. Praise and blame, on the other hand, do imply moral culpability, and so the fully rational agent could only adopt these attitudes for consequentialist purposes.

The key point to take from this, I believe, is that the perceived tension between the absence of desert on the one hand, and our holding of reactive attitudes that imply that desert is possible on the other, arises from the false assumption that we can be fully rational beings. While Smilansky says that we must maintain a motivated obscurity regarding the absence of ultimate responsibility, I say that we cannot help but (erroneously) see each other as deserving of moral responsibility ascriptions. Nevertheless, these erroneous judgements are permissible, since praising and blaming in this way can be justified on consequentialist grounds. A perfectly rational being, by contrast, would have correspondingly perfect insight into the fact that these moral responsibility ascriptions are not deserved; but they would not, for all that, choose to abandon all attitudes of praise and blame, since they would be able to appreciate at the same time the beneficial consequences of retaining our system of moral responsibility.

There is much more to say about what a fully worked-out consequentialist justification of moral responsibility should look like. I think that perhaps the most difficult aspect of continuing in this task will be to persuade doubters that a consequentialist theory is actually able to capture what it is we are doing—or should be doing—when we hold each other morally responsible. Scanlon's concern that the

beneficial consequences of expressing disapproval "depends on too many factors other than the nature of the act in question" is doubtless one that many others share, and more needs to be done to convince these sceptics that this concern can be surmounted.[6]

Part of this case for the necessity of adopting a consequentialist approach to moral responsibility will come from continuing to state the case as assuredly as possible *against* belief in desert, since if desert were possible then we would have no need to seek an alternative justification for our moral responsibility norms. Of course, persuading people of the impossibility of desert will not by itself suffice to persuade them of the wisdom of embracing a consequentialist approach. Both Scanlon and Bennett, whose objections to consequentialism were rehearsed in Chapter 7, show some sympathy towards the view that desert is impossible (although ultimately neither of them fully embrace this position). Scanlon declares: "our attitude toward those who suffer or are blamed should not be "You asked for this" but rather "There but for the grace of God go I.""[7] Bennett, meanwhile, suggests that "if a person is as God or Nature made him, and if how he is determines what he does, then it is [to quote Bernard Williams] 'in some ultimate sense hideously unfair' that he should be blamed for what he does."[8] So, sympathy for the case against ultimate desert does not automatically translate into sympathy for a consequentialist approach to moral responsibility.

As for the positive arguments for consequentialism, more can be done to make the case that holding one another morally responsible when and only when the two conditions identified are met really would bring beneficial consequences. Additionally, more work is needed to persuade sceptics that there is a good fit between our moral responsibility practices as they stand and a consequentialist justification of these practices. Only by tackling these issues can we help to ensure the credibility of the theory.

Finally, besides the possibility of providing a fuller consequentialist account, what other avenues for exploration are there following this thesis? There is certainly room for closer examination of arguments touched on in Chapter 2 concerning the

[6] Scanlon (1995), p. 48.
[7] Scanlon (1995), p. 294.
[8] Bennett (1980: p. 25) quoting Williams (1973), p. 228.

metaphysical implications of quantum theory. The battle for supremacy between "hidden variable" theories such as Bohm's and the "standard interpretation," which hinges on a debate as to which provides the best model of what goes on at the quantum level, is very much ongoing. I took a stand on that debate by siding with those who say that "hidden variable" theories are to be preferred, but there is no doubt that these arguments could be examined in greater depth. Another avenue of exploration—one that that I passed over on account of my intuition in favour of the origination condition and which, for that reason, would appeal to those who do not share this intuition—would be to survey the various traditional compatibilist responses to the challenge of origination. That is, since non-consequentialist compatibilists must deny the need for origination, it would be interesting to see how they justify this denial, and interesting also to compare and contrast the conditions under which they propose we can be morally responsible.

BIBLIOGRAPHY

Adams, Robert Merrihew (1994) *Leibniz: Determinist, Theist, Idealist*, Oxford: Oxford University Press.

Alloy Lauren B., and Lyn Y. Abramson (1979) "Judgment of Contingency in Depressed and Non-Depressed Students: Sadder but Wiser?" *American Psychological Association*, 108: 441-85.

Arneson, Richard (2003) "The Smart Theory of Moral Responsibility," in *Desert and Justice*, ed. Serena Olsaretti, Oxford: Clarendon Press.

Aristotle (1939) *De Caelo*, trans. W.K.C. Guthrie, London: William Heinemann.

_____ (1994) *Posterior Analytics*, trans. J. Barnes, Oxford: Clarendon Press.

Ayer, A. J. (1954) *Philosophical Essays*, London: Macmillan.

Beebee, Helen (2003) "Local Miracle Compatibilism," *Noûs*, 37: 258-77.

Bell, J. S. (1987) *Speakable and Unspeakable in Quantum Mechanics: Collected Papers on Quantum Philosophy*, Cambridge, Cambridge University Press.

Belot, Gordon (2001) "The Principle of Sufficient Reason," *The Journal of Philosophy*, 98: 55-74.

Bennett, Jonathan (1984) *A Study of Spinoza's Ethics*, Indiana: Hackett Publishing Company.

_____ (1980) "Accountability," in *Philosophical Subjects: Essays Presented to P. F. Strawson*, ed. Zak van Straaten, Oxford: Clarendon Press.

Bishop, Robert C. (2002) "Chaos, Indeterminism, and Free Will," in *The Oxford Handbook of Free Will*, ed. Robert Kane, Oxford: Oxford University Press.

Borges, Jorge Luis (1970) "The Garden of Forking Paths," in *Labyrinths: Selected Stories and Other Writings*, Harmondsworth: Penguin.

Cicero, Marcus Tullius (1991) *On Fate*, trans. R.W. Sharples, Warminster, England: Aris and Phillips.

Clarke, Randolph (2005) "On an Argument for the Impossibility of Moral Responsibility," *Midwest Studies in Philosophy*, 29: 13-24.

Clarke, Samuel (1998) *A Demonstration of the Being and Attributes of God, and Other Writings*, ed. E. Vailati, Cambridge: Cambridge University Press.

Cushing, James T. (2000) "Bohmian Insights into Quantum Chaos," *Philosophy of Science*, 67: 430-45.

Daumer, M, D. Dürr, S. Goldstein, and N. Zhangi (1997) "On the Quantum Probability Flux Through Surfaces," *Journal of Statistical Physics*, 88: 967-77.

Deutsch, David (1997) *The Fabric of Reality*, London: Allen Lane.

Della Rocca, Michael (2010) "PSR," *Philosophers' Imprint*, 10: 1-13.#

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

Double, Richard (2004) "The Ethical Advantages of Free Will Subjectivism," *Philosophy and Phenomenological Research*, 69: 411-22.

Dürr, Detlef, Sheldon Goldstein, Stefan Teufel, and Nina Zhangi (2000) "Scattering Theory from Microscopic First Principles," *Physica A*, 279: 416-31.

_____, Sheldon Goldstein, and Nino Zanghi (1992) "Quantum Chaos, Classical Randomness, and Bohmian Mechanics," *Journal of Statistical Physics*, 68: 259-70.

Ekstrom, Laura (2002) "Libertarianism and Frankfurt-style Cases," in *The Oxford Handbook of Free Will*, ed. Robert Kane, Oxford: Oxford University Press.

_____ (2000) *Free Will: A Philosophical Study*, Boulder, Colorado: Westview Press.

Feltz, Adam, Edward T. Cokely, Thomas Nadelhoffer (2009) "Natural Compatibilism versus Natural Incompatibilism: Back to the Drawing Board," *Mind & Language*, 24: 1-23.

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

Fischer, John Martin (2004) "Responsibility and Manipulation," *Journal of Ethics*, 8: 145-77.

Fischer, John Martin, Mark Ravizza (1998) *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge, Cambridge University Press.

Gale, Richard M., and Alexander R. Pruss (1999) "A New Cosmological Argument," *Religious Studies*, 35: 461-76.

Ginet, Carl (2007) "An Action can be Both Uncaused and Up to the Agent," in *Intentionality, Deliberation and Autonomy: The Action-theoretic Basis of Practical Philosophy*, eds. Christoph Lumer and Sandro Nannini, Aldershot: Ashgate Publishing.

_____ (1995) "Freedom, Responsibility, and Agency," *Journal of Ethics*, 1: 85-98.

_____ (1990) *On Action*, Cambridge, Cambridge University Press.

_____ (1966) "Might We Have No Choice?" in *Freedom and Determinism*, ed. Keith Lehrer, New York: Random House.

Gurr, John Edwin (1959) *The Principle of Sufficient Reason in Some Scholastic Systems*, Milwaukee: Marquette University Press.

Haji, Ishtiyaque (1998) *Moral Appraisability: Puzzles, Proposals, and Perplexities*, New York: Oxford University Press.

Heisenberg, W (1927) "Über den Anschaulichen Inhalt der Quantentheoretischen Kinematik und Mechanik," *Zeitschrift für Physik*, 43.

Hick, John (1964) *The Existence of God*, London: Collier-Macmillan.

Hodgson, David (2002) "Quantum Physics, Consciousness, and Free Will," in *The Oxford Handbook of Free Will*, ed. Robert Kane, Oxford: Oxford University Press.

Hume, David (2007) *An Enquiry Concerning Human Understanding: And Other Writings*, Cambridge, Cambridge University Press.

_____ (1972) *A Treatise of Human Nature*, London: Fontana.

_____ (1998) "An Enquiry Concerning the Principles of Morals," Oxford, Oxford University Press.

_____ (1965) *Hume's Ethical Writings*, ed. Alasdair C. MacIntyre, New York: Collier Books.

Hurley, Susan L. (2000) "Is Responsibility Essentially Impossible?" *Philosophical Studies*, 99: 229-68.

Jackson, Frank (1991) *Conditionals*, Oxford: Oxford University Press.

Kane, Robert (1999) "Responsibility, Luck, and Chance: Reflections on Free Will and Indeterminism," *Journal of Philosophy*, 96: 217–240.

_____ (1996) *The Significance of Free Will*, Oxford: Oxford University Press.

Kant, Immanuel (2004) *Prolegomena to Any Future Metaphysics*, trans. Gary C. Hatfield, Cambridge: Cambridge University Press.

_____ (1997) *Lectures on Metaphysics*, trans. S. Naragon, Cambridge: Cambridge University Press.

Katz, B.D. (1997) "The Cosmological Argument Without the Principle of Sufficient Reason," *Faith and Philosophy*, 14: 62-70.

Lamb, James W. (1977) "On a Proof of Incompatibilism," Philosophical Review, 86: 20-35.

Leavens, C. Richard (1996) "The "Tunneling-Time Problem for Electrons," *Boston Studies in the Philosophy of Science*, 184: 111-29.

Leftow, B. (1988) "A Modal Cosmological Argument," *International Journal for Philosophy of Religion*, 24: 159-88.

Leibniz, G.W.F. (1991) *Discourse on Metaphysics and Other Essays (On the Ultimate Origination of Things; Preface to the New Essays; The Monadology)*, trans. R. Ariew and D. Garber, Indiana: Hackett Publishing Company.

_____ (1989) *Philosophical Essays*, trans. R. Ariew and D. Garber, Indiana: Hackett Publishing Company.

_____ (1986) *Sämtliche Schriften und Briefe*, Berlin: Akademie-Verlag.

_____ (1956a) *Philosophical Papers and Letters*, trans. Leroy E. Loemaker, Chicago: University of Chicago Press.

Leibniz G.W.F., and S. Clarke (1956b) *The Leibniz-Clarke Correspondence*, ed. H.G. Alexander, Manchester: Manchester University Press.

Leslie, John (2001) *Infinite Minds: A Philosophical Cosmology*, Oxford: Clarendon Press.

Lewis, David (1986) *On the Plurality of Worlds*, Oxford: Basil Blackwell.

_____ (1981) "Are We Free to Break the Laws?" *Theoria*, 47: 113-21.

Lin, Martin (2007) "Spinoza's Arguments for the Existence of God," *Philosophy and Phenomenological Research*, 75: 269-97.

Longuenesse, B. "Kant's Deconstruction of the Principle of Sufficient Reason," *Harvard Review of Philosophy*, 9: 67-87.

Mackie, J. L. (1977) *Ethics: Inventing Right and Wrong*, Harmondsworth: Penguin.

Mele, Alfred R. (1999) "Kane, Luck, and the Significance of Free Will," *Philosophical Explorations: An International Journal for the Philosophy of Mind and Action*, 2: 96-104.

_____ (1995) *Autonomous Agents: from Self-Control to Autonomy*, Oxford: Oxford University Press.

Meyer, R.K. (1987) "God Exists!" *Noûs*, 21: 345-61.

McCallum, James Ramsey (1976) *Abelard's Christian Theology*, Merrick, New York: Richwood Publishing Company.

Mill, John Stuart (1872) *An Examination of Sir William Hamilton's Philosophy and of the Principal Philosophical Questions Discussed in His Writings*, London: Longmans, Green, Reader, and Dyer.

Misenheimer, L. (2008) "Predictability, Causation, and Free Will," (Unpublished Manuscript), University of California: Berkeley.

Nahmias, Eddy, Stephen G. Morris, Thomas Nadelhoffer, and Jason Turner (2006) "Is Compatibilism Intuitive?" *Philosophy and Phenomenological Research*, 73: 28-53.

_____ (2005) "Surveying Freedom: Folk Intuitions about Free Will and Moral Responsibility," *Philosophical Psychology*, 18: 561-84.

Neitzche, Friedrich (2008) *Twilight of the Idols*, Oxford: Oxford Paperbacks.

_____ (1973) *Beyond Good and Evil: Prelude to a Philosophy of the Future*, Harmondsworth: Penguin.

Nichols, Shaun, Joshua Knobe (2007) "Moral Responsibility and Determinism: The Cognitive Science of Folk Intuitions," *Noûs*, 41: 663-85.

Nozick, Robert (2012) *Anarchy, State, and Utopia*, Oxford: Blackwell.

Parfit, Derek (1984) *Reasons and Persons*, Oxford: Clarendon Press.

Parmenides (1986) *The Fragments of Parmenides*, trans. A.H. Coxon, Assen, Netherlands: Van Gorcum.

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

Plato (1971) *Timaeus and Critias*, trans. H.D.P. Lee, Harmondsworth: Penguin.

Pojman, Louis P. (2003) *Philosophy of Religion: An Anthology*, United Kingdom: Wadsworth/Thomson Learning.

Pruss, Alexander R. (2011) *The Principle of Sufficient Reason: A Reassessment*, Cambridge: Cambridge University Press.

Reichenbach, Bruce (1972) *The Cosmological Argument: A Reassessment*, Springfield, Illinois: Thomas.

Rescher, Nicholas (2000) "Optimalism and Axiological Metaphysics," *The Review of Metaphysics*, 53: 807-35.

Roskies, A. and S. Nichols (2008) "Bringing Moral Responsibility Down to Earth," *Journal of Philosophy*, 105: 371–88.

Rowe, William L. (1975) *The Cosmological Argument*, London: Princeton University Press.

Sarkissian, Hagop, Amita Chatterjee, Felipe de Brigard, Joshua Knobe, Shaun Nichols, and Smita Sirker (2010) "Is Belief in Free Will a Cultural Universal?" *Mind & Language*, 25: 346-58.

Scanlon, T. M. (1995) "The Significance of Choice," in *Equal Freedom: Selected Tanner Lectures on Human Values*, ed. Stephen Darwall, Ann Arbor: University of Michigan Press.

Schopenhauer, Arthur (1974) *On the Fourfold Root of the Principle of Sufficient Reason*, trans. E.F.J. Payne, La Salle, Illinois: Open Court.

Schrödinger, E. (1936) "The Present Status of Quantum Mechanics," *Die Naturwissenschaften*, trans. John D. Trimmer in the (1980) *Proceedings of the American Philosophical Society*, 124: 323-38

Slote, Michael (1982) "Selective Necessity and the Free-Will Problem," *The Journal of Philosophy*, 79: 5-24.

Smart, J.J.C. (1973) "An Outline of a System of Utilitarian Ethics," in *Utilitarianism: For and Against*, eds. J. J. C. Smart and Bernard Williams, Cambridge: Cambridge University Press.

_____ (1961) "Free-Will, Praise and Blame," *Mind*, 70: 291-306.

Smilansky, Saul (2002) "Free Will, Fundamental Dualism, and the Centrality of Illusion," in *The Oxford Handbook of Free Will*, ed. Robert Kane, Oxford: Oxford University Press.

_____ (2000) *Free Will and Illusion*, Oxford: Clarendon Press.

Spinoza, Benedictus de (1992) *Ethics, Treatise on the Emendation of the Intellect, and Selected Letters*, trans. Samuel Shirley, Indiana: Hackett.

Stern, Alfred (1969) *Individuals: An Essay in Descriptive Metaphysics*, London: Methuen.

Steward, Helen (2012) *A Metaphysics for Freedom*, Oxford: Oxford University Press.

Strawson, Galen (2010) *Freedom and Belief*, Oxford: Oxford University Press.

_____ (2002) "The Bounds of Freedom," in *The Oxford Handbook of Free Will*, ed. Robert Kane, Oxford, Oxford University Press.

_____ (2000) "The Unhelpfulness of Indeterminism," *Philosophy and Phenomenological Research*, 60: 149-55.

_____ (1994) "The Impossibility of Moral Responsibility," *Philosophical Studies*, 75: 5-24.

Strawson, Peter F. (1962) "Freedom and Resentment," *Proceedings of the British Academy*, 48: 1-25, republished in ed. Gary Watson (1982) *Free Will*, Oxford: Oxford University Press, pp. 59-80.

Taylor, Richard (1963) *Metaphysics*, Englewood Cliffs, New Jersey: Prentice-Hall.

Van Inwagen, Peter (2009) *Metaphysics*, Boulder, Colorado: Westview.

_____ (2004) "Freedom to Break the Laws," *Midwest Studies in Philosophy*, 28: 334-50.

_____ (1983) *An Essay on Free Will*, Oxford: Clarendon Press.

_____ (1975) "The Incompatibility of Free Will and Determinism," *Philosophical Studies*, 27: 185-99.

Vargas, Manuel (2011) "Revisionist Accounts of Free Will: Origins, Varieties, and Challenges," in *The Oxford Handbook of Free Will: Second Edition*, ed. Robert Kane, Oxford, Oxford University Press.

_____ (2005) “The Revisionist’s Guide to Responsibility,” *Philosophical Studies*, 125: 399-429.

_____ (2004) “Libertarianism and Skepticism about Free Will: Some Arguments against Both,” *Philosophical Topics*, 32: 403-26.

Viney, Wayne, David A. Waldman, Jacqueline Barchilon (1982) “Attitudes Towards Punishment in Relation to Beliefs in Free Will and Determinism,” *Human Relations*, 35: 939-50.

Wallace, David, Simon Saunders, Jonathan Barrett, and Adrian Kent (2012) *Many Worlds? Everett, Quantum Theory, & Reality*, Oxford: Oxford University Press.

White, David E. (1979) “An Argument for God’s Existence,” *International Journal for Philosophy of Religion*, 10: 101-15.

Wiggins, David (1973) “Towards a Reasonable Libertarianism,” in *Essays on Freedom of Action*, ed. Ted Honderich, London: Routledge.

Williams, Bernard (1973) *Problems of the Self*, Cambridge: Cambridge University Press.

Wolff, Christian von (1736) *Philosophia Prima Sive Ontologia: Methodo Scientifica Pertractata, Qua Omnis Cognitionis Humanae Principia Continentur*, Francofurti Et Lipsiae: Renger.

Woolfolk, Robert L., John M. Doris, and John M. Darley (2006) “Identification, Situational Constraint, and Social Cognition: Studies in the Attribution of Moral Responsibility,” *Cognition*, 100: 283-301.