FATALISM AND THE REALITY OF THE FUTURE

by

Robert Lawrence Nichols

A thesis submitted for the degree of Doctor of Philosophy of the Australian National University, June 1997
Declaration

This thesis is my own work. All sources used have been acknowledged.

Robert Lawrence Nichols

Acknowledgements

I would like to thank Brian Garrett, Paul Thom, Peter Roeper, David Braddon-Mitchell for their generous help and criticism. For their assistance and helpful comments, I also thank Michael Tooley, Graham Nerlich, Dominic Hyde, Andrew Gleeson, Joe Mintoff, Arthur Witherall, Jack Smart and Chris Falzon.

I am also indebted to my family for their support, and especially to Lee Galloway for her constant support and encouragement.
Abstract

This thesis attempts a defence of fatalism through an examination of the main arguments for and against it. I argue that the issue of fatalism has significant connections with the two main current theories of time. Many have thought that the tenseless, or so-called 'B', theory of time entails fatalism. I contend that this claim is in fact justified, but that we are not entitled to reject the theory on this basis, since the argument for fatalism, when properly understood, is itself sound. I further contend that it is as a direct consequence of the B-theory's commitment to the full and determinate reality of the future that the fatalist conclusion follows.

The thesis is divided into four main sections. In the Part One, I discuss various misconceptions concerning fatalism, outlining the doctrine's actual and imagined commitments, before proceeding to a discussion of the two theories of time, followed lastly by an assessment of the claim that the B-theory is fatalistic. In Part Two, I examine both the oldest argument for truth fatalism as well as Richard Taylor's modern condition version. In the course of this examination, I advance the case for ontological fatalism. This designation is intended to capture the fact that fatalistic constraint follows from ontological rather than semantic or causal facts. In Part Three, I examine three cases where fatalism has been thought to figure. In the course of an extended examination of the principal paradox of time travel, it is argued not only that fatalism provides a resolution of the paradoxical situation, but further, that the lessons learned here can be applied to our ordinary situation. Finally, in Part Four, I show how the truth of fatalism need not conflict with the acknowledged reality of the phenomenon of human deliberation.
PART 1 - FATALISM & TIME

Chapter 1: Introduction - The Nature and Scope of Fatalism

1.1 The Peremptory Dismissal of Fatalism
1.2 A Definition of Fatalism
1.3 What Fatalism is and is not
1.4 The Role of the Future

Chapter 2: Theories of Time

2.1 Ordinary Ideas about Time
2.2 Tensed and Tenseless Views of Time
2.3 Differing views of the Past and Future
2.4 The New Theory of Time
2.5 Problems of the A- and B-Theories
2.6 The Relativistic Woes of the A-theory

Chapter 3: The Relationship between Fatalism and Theories of Time

3.1 Fatalism and the A-theory
3.2 Fatalism and the B-theory
3.3 The B-theory Does Entail Fatalism

PART 2 - NAVAL LOGIC, PAST & FUTURE

Chapter 4: Aristotle's Discussion of Fatalism

4.1 The Traditional Argument for Fatalism
4.2 Competing Interpretations
4.3 The Structure of De Interpretatione 9
4.4 The Two Arguments
4.5 Basic Version of the Argument for Truth Fatalism
4.6 Strategy (a): Restricting Bivalence
Chapter 8: Newcomb's Problem

8.1 An Outline of the Problem
8.2 Decision Theory Principles
8.3 Some Possible Responses
8.4 Explaining the Predictor's Success

PART 4 - CONCLUSION

Chapter 9: Contingency, Efficacy & Deliberation

9.1 The Consequences of Fatalism
9.2 Fatalist Deliberation
9.3 The Accusation of Insincerity
9.4 The Pointlessness of Deliberation about the Future
9.5 Concluding Remarks

Bibliography
PART ONE

FATALISM AND TIME
CHAPTER 1

INTRODUCTION - THE SCOPE AND NATURE OF FATALISM

Stärker als die Zeit war das Schicksal, und in ihm lag der Zeiten letztes Geheimnis verborgen.
(Stronger than time is fate, in which the final secret of time lay hidden.)
Hermann Broch
The Death of Virgil

Ducunt fata volentem, nolentem trahunt.
(Fate leads the willing, and drags the unwilling.)
Seneca

Fatalism is intimately bound up with time. The failure to take fatalism seriously is a direct result of neglecting these profound connections. Mark Bernstein, in one of the more thoughtful articles on fatalism, has observed that, in the absence of compelling theories about truth and time, both the fatalist and the anti-fatalist are condemned to endless rounds of reciprocal question-begging. While no one since Aristotle has ever doubted the relevance of truth to the question of fatalism, the relationship between fatalism and time has been less successfully plumbed. What treatment there has been of this issue has usually revolved merely around the fatalistic implications of the two main contending theories of time. However the connection is far more profound. It is not just a question as to whether one or the other of these theories is fatalistic: rather what we shall discover is that the very ontological commitment of one of these theories with respect to the future installs the most inescapable argument for the truth of fatalism.

1.1 The Peremptory Dismissal of Fatalism

'It is certainly not the least charm of a theory that it is refutable: it is with precisely this charm that it entices subtler minds.' 1 Friedrich Nietzsche, when he made this observation in Section 18 of Beyond Good and Evil, was not talking about fatalism. (Indeed, the comments were made in reference to the antithetical doctrine of 'free will'.) 2 However many other philosophers have thought of fatalism, not only that it is eminently refutable, but that it has in fact been well and truly refuted. Gilbert Ryle

---

1 Nietzsche (1973: 29).
2 'It seems that the hundred times refuted theory of "free will" owes its continued existence to this charm alone—: again and again there comes along someone who feels he is strong enough to refute it' Nietzsche (1973: 29).
suggests that it is unworthy of prolonged attention: 'No philosopher of the first or second rank has defended fatalism or been at great pains to attack it.' Undaunted, some philosophers have come forward to defend it; Richard Taylor is the most notable contemporary example. This is, however, a thankless task: the prevailing attitude has tended to be not merely critical, but indeed peremptory and dismissive. For William Lane Craig, '[t]he argument for fatalism must be fallacious, and we need only find the error'. Margery Bedford Naylor agrees, passing the following judgment on Taylor's long-standing advocacy of the doctrine:

We might find his arguments compelling if we were not so certain that fatalism must be false. It is much easier, however, to see that something must be wrong with them than it is to say what is wrong with them.

This difficulty has nonetheless dissuaded few from trying to articulate where it is that fatalism goes astray. I will perforce examine most of these criticisms at various points in what follows.

This thesis is born out of the conviction that at least some forms of both the traditional and contemporary arguments in favour of fatalism are, when correctly interpreted, sound. I am also convinced that the whole debate is most usefully viewed in the context of the great twentieth century debate between the tensed and the tenseless views of time. In particular, I claim that the issue of the ontological status of the future is crucial to the assessment of the adequacy of fatalist arguments. A principal contention will be that one of these theories, viz., the tenseless view, does, as it has so often been accused of by its detractors, have fatalistic implications, but that since fatalism is in fact true, this consequence is not a liability but a boon.

In this introductory chapter, I will examine what fatalism is and what it is not. In the following chapter I outline the two contending theories of time, and give at least an indication of how these debates intersect. Subsequent chapters will focus on (i) the question of whether or not the tenseless view of time entails fatalism, (ii) the standard arguments, from Aristotle to Taylor, for the truth of fatalism, (iii) the role played by fatalist considerations in some recent and seemingly intractable debates surrounding time travel, backwards causation, and Newcomb's choice problems, and

---


4 Craig (1991: 43).

5 Naylor (1980: 49).
finally (iv) a consideration of some of the reasons why there appears to have been such unreflective resistance to fatalism based on its consequences, real and imagined.

1.2 A Definition of Fatalism?

Fatalism may be loosely defined as the claim that everything that happens is necessary, or happens of necessity. According to Mark Bernstein, '[f]atalism tells us that whatever happens had to happen.' It is clear that 'had to' here refers to some sort of necessity and it will have to be spelled out precisely what sense of 'necessary' this is. But if whatever will happen must happen, then I (or for that matter any agent or influence) cannot prevent whatever happens from happening. It is ordinarily taken to be the case that we, as agents, possess the power to choose our actions in the sense that at least some of our actions are such that we could act otherwise than the way we do act. Assume I perform action A. Anti-fatalists, who believe in the reality of human freedom, claim that the following proposition is true:

\[(A) \quad \text{It is within my power to do A, and it is within my power to do } \neg A.\]

That is to say, I performed A, but I could have performed \(\neg A\), had I so chosen. The denial that propositions like (A) are ever true is just what it is that renders one a fatalist. The fatalist holds that all events are unavoidable; the anti-fatalist holds that some events are avoidable. We can define philosophical fatalism then as the following view:

\[(F) \quad \text{Whatever happens is, and always was, unavoidable.}\]

The fatalist denies that the future is a realm of possibilities from which we may choose some and pass by others. We therefore lack the ability to perform any actions other that those we do. And the salient point about this lack is that it derives from 'logical considerations about truth, rather than from empirical considerations

---

6 Compare the OED (s.v. "Fatalism"): "In early use not distinguished from 'the doctrine of necessity', i.e., the doctrine that all events take place in accordance with unvarying laws of causation. In strict etymological propriety, and in the best modern usage, it is restricted to the view which regards events as predetermined by an arbitrary decree." It derives from fatum, lit. "that which has been spoken", a sentence or doom of the gods. This OED sense is certainly not the sense I shall be concerned with. Rather it is the sense of fatum as something 'decreed' in speech or writing that is responsible for the oft-found conceit of a Book of Nature or an Ideal Chronicle, containing the set of all true propositions in a given domain.
concerning the operations of either the world or human nature'. The important idea is not just that no agent is able to act otherwise than he or she in fact does, but that they are unable to act otherwise for logical or conceptual reasons having to do with truth. Consequently the argument for fatalism is often couched in terms of the truth or falsity of propositions about the future and/or logical considerations such as the Law of the Excluded Middle. (I will later issue a qualification about the role played by truth here, but let this stand for now.) Logical fatalism is hence often characterised alternatively as 'truth' fatalism.

This is (tentatively) what I will understand by fatalism. Naturally a great many other definitions have been offered—often by opponents—not all of which coincide with this one. The variations are however often instructive in their own right. While it is important for the defender of fatalism to warn against misleading and inappropriate attributions and implications, one must be careful here: the fatalist needs to recommend a claim that is weak enough to be plausible, yet sufficiently robust to be interesting. The ultimate plausibility of this truth fatalism has often been impugned. I believe much of this implausibility can be side-stepped if fatalism is couched in ontological rather than alethic terms. It is ontologically grounded since its lesson of unavoidability arises out of the determinately real existence of actual states of affairs. I hope to show that fatalism thus conceived passes the robustness requirement.

1.3 What Fatalism Is and Is Not.

What then is fatalism? The most important thing to realise right at the start is that fatalism is an ontological thesis. It is a view about the events that go to make up the world, and in particular a view about the manner and nature of their occurrence. Fatalism views the occurrence of all events—past, present or future—as partaking in the necessity that commonsense ascribes to past events. The past is closed in the sense that no one now has any ability to modify or 'rearrange the furniture of the past', to use Mark Bernstein's felicitous phrase. The fatalist sees the future as just as closed as the past in that we have no power either to modify the events that exist there. Fatalist necessity then is cut from the same cloth as the necessity of the past.

It is often said that while we are all fatalists about the past, in that we now believe that this realm is beyond our control, we remain committed to the idea that

---

7 (Bernstein 1990: 270).
8 Bernstein (1992: 8).
fatalism about the future is false, in the sense that the future may be affected by us more or less as we choose. Within, of course, certain physical limits—I cannot, for example, choose to stop falling halfway between the aeroplane I have just jumped out of and the ground. This asymmetry mirrors some temporally based ones we will examine in the next chapter. I intend to argue that, contrary to this common attitude, fatalism is no less true concerning future events than past ones.

Some preliminary misconceptions about what fatalism is and isn't, and concerning what it entails and what it doesn't, need to be addressed. Fatalism as I understand it proceeds from no special assumptions about causal determinism nor about the existence of an omniscient, foreknowing God. Neither does fatalist necessity equate with logical necessity, or entail that all truths are necessary truths. Nor does its denial of avoidability reduce to the empty truism that what will be will be. Most importantly, it is not committed to the inefficaciousness of all human action. I will say a little about each of these in turn (spending somewhat more time on the last since it embodies such a widespread misconception).

Logical Necessity

We saw above that fatalist necessity is of a piece with the so-called necessity of the past. It is important to realise what this necessity is not. The concept of necessity appealed to by the fatalist is not logical necessity. It cannot, for example, be the idea that if it is true that I will die tomorrow, then it is logically necessary that I will die tomorrow. To take the fatalist who says that I cannot prevent what will happen to be saying that the prevention of this future event is logically impossible is to attribute an absurd doctrine to him. We will see that what I can do, according to the fatalist, is not just what is logically possible but rather that which is within my power. And according to the fatalist the only things that are within my power, are those things I in fact do.

Necessary Truth

The fatalist need not deny the widespread use of possible worlds discourse in modern philosophy, which would have it that there are some possible worlds in which I die tomorrow and others in which I survive. Fatalism is a thesis about the actual world, but quite compatible with talk about other logically possible worlds. This is why even though Felt (1992) starts out well in defining fatalism as 'the view that whatever happens, happens inevitably and could not have happened otherwise', he goes wrong immediately thereafter when he continues: 'It is reductively the view that the actual
and the possible coincide, for if nothing could happen otherwise than as it does, then the actual exhausts the possible.' 9

Still less can the fatalist's contention be that there is only one possible world, namely, the actual world, or that there exist no unrealised possibilities, which is, as van Inwagen observes 'a doctrine equivalent to the doctrine that a proposition is true if and only if it is necessarily true.' 10 It is surely not a necessary truth that I die tomorrow even if it is true that I will. Nor does it seem plausible to deny that I can talk about the (admittedly unrealised) possibility that I survive till next week.

*The emptiness of 'avoidable' ['What will be, will be']*

J.J.C. Smart finds fault with the fatalist's use of 'unavoidable'. He complains about building the assumption that a given event E occurs into the contextually agreed background assumptions for a given event, implying that this renders the sense of 'unavoidable' to which the fatalist appeals trivial and utterly uninteresting:

If the fatalist says that event E is fated and then includes in background assumptions the statement that event E occurs he has implicitly adopted an empty sense of 'avoidable' which should not worry us any more than 'What will be will be' should worry us. If this is fatalism it is a totally empty doctrine.11

Were there no good reason for so building them in, this might indeed be objectionable. However, this is where the B-theoretic commitment to the determinate reality of the future is relevant. Given this commitment, E either is or isn't there in these background assumptions (or rather, either does or does not belong there). If it is there, then the event is unavoidable in an interesting sense. In that case, you can't choose E v ¬E as you choose, though you can choose E or you can choose ¬E (just not either one as you please).

Hence it is wrong to claim that fatalism is objectionable because it is just a matter of 'whatever will be, will be'. Rather, fatalism trades on the hitherto unsuspected implications of the ontological commitments of the B-theory: if E will be the case then E is unavoidable, and if ¬E, then it is unavoidable. Of course, we may not ever know which is (or will be) the case in each instance. But that we may

---

11 Smart (1989: 154), but see the extended discussion on pp. 149-54.
not (even cannot possibly) know which unavoidable event this will be, does not diminish the fact that it is a more interesting and substantial claim than the merely tautologous 'what will be, will be'. The really existing events that comprise the world are non-negotiable features of the ontological landscape.

Logical versus theological fatalism

Some commentators see these two doctrines as entirely separate. Many in the long line of commentators on the problem of divine foreknowledge believe that while the problem of logical fatalism is easily solved since the argument for it is demonstrably unsound—either in virtue of false premises or invalidity (or both)—the problem of theological fatalism is a far thornier one. Others, such as Susan Haack (1974b) and William Lane Craig (1991), believe that theological fatalism ultimately depends on logical fatalism.12 If this is the case, then once you have disposed of logical fatalism, no more work need be done since the defect in the argument for logical fatalism will infect the argument for theological fatalism as well.

Nelson Pike has referred to the following as the 'controlling premise' in the foreknowledge argument for (theological) fatalism:

(CP) If God existed and believed at t1 that Jones would do A at t2 and if Jones has power at t2 to refrain from doing A, then either (1) Jones has power at t2 to act in such a way that, were he to act in that way, God would have held a false belief at t1; or (2) Jones has power at t2 to act in such a way that, were he to act in that way, God would have existed but would not have held the belief he held at t1; or (3) Jones has power at t2 to act in such a way that, were he to act in that way, God would not have existed at t1.

For Pike, none of these consequences 'can be accepted'.13 The conclusion that it is not within anyone's power to do anything other than what they actually do is said to follow from the existence, omniscience and eternity of God. Faced with an argument

12 For Haack, the argument for theological fatalism is just 'a needlessly (and confusingly) elaborated version of Aristotle's De Interpretatione 9 version and worrying about omniscient deities merely deflects attention from the real issue of the truth of future contingent propositions (1974b: 158). For Craig 'the problem of theological fatalism is reducible to a statement of purely logical fatalism involving no reference to God or any knower' (1991: 32).

for fatalism based on this premise, many would prefer to deny the existence, or perhaps the omniscience of God. Any or all of these would seem to be attractive candidates for jettisoning on the part of those disinclined to accept fatalism. While this type of manoeuvre may be a live option in the theological case, where the existence of the deity may after all be in any case improbable, it will not do in the logical case, since everyone is apt to agree that truth, in some form or other, is real. I therefore believe that logical, or truth, fatalism is inherently more interesting, not just because it avoids such problematical ontological commitments, but because there would appear to be an uncontentious sense in which 'truth' and 'state of affairs', which are the essential elements of this form of fatalism, manifestly do exist. In any case, whether or not the theological variety is reducible to the logical one, I will be concerned almost exclusively with the latter, treating the former tangentially, if at all. There are problems enough, even in this more restricted patch.

Fatalism versus Causal Determinism

It is absolutely crucial to understand that the claim embodied in (F) has nothing whatsoever to do with the issue of determinism or indeterminism. In particular, it is not based in any way on, nor equivalent to, nor does it entail causal determinism. According to William James, '[determinism] professes that those parts of the universe already laid down absolutely appoint and decree what the other parts shall be. The future has no ambiguous possibilities hidden in its womb: the part we call the present is compatible with only one totality.'\(^{14}\) The crucial feature here is the idea that some 'parts' of the world causally determine future 'parts'.

The fatalist's claim is merely that a given event is unavoidable. He says nothing at all about how that event came about. It may be that the operation of some law of nature or the occurrence of some prior event renders the given event unavoidable; for example, if the event in question is an eclipse, or my death, in the event of my having leapt from a sufficiently tall building. But the fatalist's claim would still hold even in the absence, whether contingently or in principle, of some identifiable cause. Fatalism is entirely consistent even with the sorts of indeterminism currently thought to be suggested by quantum indeterminacy. A subatomic event such as the decay of a given particle is no less a fixed and unavoidable feature of reality for not being a strictly causally determined event. And it is this quality that renders the event unavoidable. So while fatalism, as we shall

\(^{14}\) Quoted in McCall (1994: 12).
soon see, can accommodate the causal determination of events, there is no essential element of fatalism that requires that they have these antecedents.

What, then, is the source of the necessity to which the fatalist is committed? I will argue that the necessity of future events results directly from the fact that the (future) events in question will occur. But if this is the case, haven't we severed all causal links between an agent's actions and the events which one would have thought result from them? For if they are necessary simply in virtue of their occurring, then how could my actions be said to have brought them about? The answer lies in the fact that these two doctrines are entirely separate, and we will do well to keep this in mind. It is in virtue of their being separate that the two views are consistent. Fatalism does not essentially involve causal determinism, nor does it expressly exclude it. The two can quite easily go hand in hand. Indeed it will often be the case that the actual existence of the future event (which is unavoidably located up there in the future) is directly attributable to some causal input from me.

The inefficacy of action

Contrary to widespread opinion, fatalism does not render all human action inefficacious. It is often said that the fatalist is committed to the view that what happens is not up to us, that 'one cannot do anything at all about the future'. The feeling is that the fatalist objectionably and implausibly is committed to the view that human actions are causally cut off from the future. Andros Loizou offers this view of fatalism: 'The substance of the fatalist thesis is thus simply that the future is in some sense there as an impending presence, and _that as a consequence no present action or non-action can in any way determine or shape the future_.' This is a complete misrepresentation of the fatalist's position, but its prevalence may explain

---

15 Holdcroft (1973: 303). Bertman (1976: 481) identifies fatalism with this claim. Richard Gaskin writes that 'the truth of a statement bearing on the future entails that there is nothing anyone can do to affect the obtaining or otherwise of the relevant state of affairs' (1995: 12). This is somewhat ambiguous: if he means to imply that fatalists believe that you can't bring about either state of affairs, more or less at will, then that's one thing; but if he means that they can't affect the actually existing state of affairs at all, then he claims too much, since it may only exist because of the agent's action of bringing it about. Cf. Jan Faye: 'We shall define fatalism as the doctrine that entails, minimally, that nothing that happens could have been otherwise, and that no person could therefore ever do anything other than what he or she actually does. It follows from this that if fatalism were true, then it would be futile to take precautions' (1989: 107).

much of the general concern over the doctrine. Fatalism is a threat because appears to imply that all our action may turn out to be mere sham and our deliberations futile. We place great store in playing an efficacious role in bringing about our future. It explains why many feel a metaphysical attraction to views of time that cater for an open future.

A measure of how widespread this misconception is can be gained when we consider that even philosophers sympathetic to fatalism make this mistake. Mark Bernstein writes 'that the doctrine proclaims our impotence concerning all events' and admits that it is 'quite unsettling' that 'we can have no abilities to affect the constitution of the world.' Even a careful fatalist like Richard Taylor sometimes falls into the trap of formulating his doctrine in ways that might suggest to the unwary that agents lack causal input if fatalism is true: 'A fatalist, then, is someone who believes that whatever happens is and always was unavoidable. He thinks it is not up to him what will happen a thousand years hence, next year, tomorrow, or the very next moment.' What Taylor means by 'not up to him' is not as clear as it should be. Nevertheless what is clear is that it doesn't follow from an event's unavoidability that an agent can have no effect whatsoever on that event's being the one that is unavoidable. Taylor's philosophical invention, Osmo, who reads of his death in the book that contains the set of all true propositions about his life, may be fated to die, but he certainly causally affects his future—indeed he fulfils it by causing the very crash in which he perishes. (That this is the result of his foreknowledge about his fate is just an epistemological flourish that is ontologically irrelevant to the set of events in which he is enmeshed. It adds irony but leaves untouched the causal and ontological features of the situation. Other instances of this unintended self-fulfilment abound, often in literature. For example, Oedipus fulfils his destiny precisely in his efforts to avoid it, and in J.B. Priestly's play Sheppy, the character 'Death' tells the story of a servant who encounters him accidentally in Bagdad and, taking fright, tries to escape by fleeing to the town of Samarra—precisely where Death had expected all along to meet him.) Needless to say the point being made here about the efficaciousness of one's actions of course in no way relies essentially on such dramatic scenarios.

19 Taylor (198(): 334-6).
20 As quoted in van Inwagen (1983: 24).
This misrepresentation about causal discontinuity, motivated by the desire to stigmatise the fatalist, lies at the root of one of the oldest and least persuasive arguments against fatalism, the so-called Idle Argument. Jan Faye summarises the main idea when he says that '[i]t follows from this [viz., that no one could ever do anything other than what they actually do] that if fatalism were true, then it would be futile to take precautions'. If the fatalist is right in holding that all action is pointless, then we might as well remain idle. The argument was first explicitly formulated by Cicero in his De Fato:

For they argue as follows: 'If it is fated for you to recover from this disease, then you will recover, whether you call in a doctor or not; [29] similarly, if it is fated for you not to recover from this disease, then you will not recover, whether you call in a doctor or do not. But one or the other is fated; so there is no point in calling in a doctor.'

Our proposed actions are alleged to be either superfluous or futile. This type of reasoning often surfaces in folk thought on fate. A soldier reasons that either the bullet has his name on it or it doesn't and thus recklessly exposes himself to fire. The smoker believes he might as well smoke as not since he will either get cancer or he won't. Shakespeare's Henry V upbraids Westmoreland for expressing a desire for reinforcements on the eve of battle:

Westmoreland. O that we now had here
But one ten thousand of those men in England
That do no work today!
King. What's he that wishes so?
My cousin Westmoreland? No, my fair cousin.
If we are marked to die, we are enow
To do our country loss; and if to live,
The fewer men, the greater share of honor.
God's will! I pray thee wish not one man more.
(Henry the Fifth, IV iii 16-23)

Westmoreland's wish was presumably based on an understandable desire to have on board an increment of men causally sufficient to tip the scales in favour of victory. Henry's point would seem to be that whether they fight alone or with reinforcements is irrelevant to whether or not they are victorious. Despite Henry's deft grasp of the Law of the Excluded Middle, his 'fatalism' is merely a rhetorical sophistry.

---

In order to see how the authentic fatalist can reply to such misprisions, imagine that I am a sick person who is trying to decide whether or not there is any point in taking the medicine that has been offered. Let us say then that there are four possible outcomes (focussing just on $A =$ my action of taking the medicine and $E =$ the event of my recovery):

(i) $A \& E$ [I take the medicine and I recover]
(ii) $A \& \neg E$ [I take the medicine yet I die]
(iii) $\neg A \& E$ [I refuse the medicine but recover nonetheless]
(iv) $\neg A \& \neg E$ [I refuse the medicine and die]

One of these conjoined outcomes will obtain. Since I don't know which this will be, I can legitimately (attempt to) choose any of them. My preferred outcomes are obviously (i) or (iii). (Given that I can afford the medicine, I may be assumed to have no real preference.) If I believe—for whatever reason—that taking the medicine is more likely to bring about an outcome characterised by $E$, then I will rightly choose $A$. (Whether I can or not so choose, is another matter, but the question at the moment is just whether or not it is pointless to even attempt a choice.)

What is relevant though is that each outcome is compatible with it being true that what I do has an effect on the world. My taking the medicine may cause me to recover or may have no effect at all (it may even, through some unforeseen medical disaster, bring about my even speedier death!)—but this does not mean that it can't affect the content of reality. The sheer fact of one (or indeed both) of these pairs of conjoined events being fated does not necessitate any absence of causal link between each pair. The gist of the fatalist’s reply is thus that the following state of affairs may obtain: the event which does occur, say my recovery, is necessary but it is only necessary because of the act of calling for the doctor—so my action in doing so can hardly be called superfluous! The fatalist can claim that while an unavoidable event cannot be prevented by any of my actions (or anybody's), it may well be brought about by my action.23

The anti-fatalist who bases his stance on the idle argument is too selective in his choice of premises. The proponent of the idle argument reasons thus: if I am going to be killed in an air-raid, then I am going to be killed if I take precautions to save myself; and, if I am not going to be killed, then I am not going to be killed if I do not take precautions. Hence all precaution-taking behaviour is pointless. But

23 Cicero reports that Chrysippus talked of such events being 'co-fated' ('confatalia') in the sense that it is as fated that the doctor will be called in as that you will get well again. (1991: 30).
what of the other two premises it is possible to formulate: if I am going to be killed, then I am going to be killed if I do not take precautions, and if I am not going to be killed, then I am not going to be killed if I do take precautions. Hence taking precautions is very much to the point. Now, I don't know whether taking precautions in this particular case is likely to be efficacious or not—and it doesn't matter, since we aren't talking about whether one event actually causes another or not, but just whether illegitimately focussing on but two of these conditional statements is sufficient to dissuade me from doing something I would otherwise do. The anti-fatalist fondly supposes that restricting attention to this restricted set would have this effect—and tries to use that fact against the fatalist. In reply, the fatalist need only point to the mistake of not considering the complete set as essential. Ultimately, all four statements are probably insufficient to convince one of anything. But it is certainly the case that just selecting two of them, as if they described fully all the possible outcomes, will not do.

The main problem with the idle argument then is that it imputes to the fatalist a simplistic form of reasoning to which he is not really committed. Despite the obvious adequacy of this reply, it is surprising how often the 'idle' reasoning rears its head, and not only among the unsophisticated. But clearly fatalism can allow that human actions have effects, does not thereby threaten through any of its implications the causal continuity of the past and present with the future. Indeed, as I have acknowledged, it will often be the case that the event in question only comes about as a direct consequence of these very human actions.

1.4 The Role of the Future

I have sketched what the fatalist is and isn't committed to. But what about the anti-fatalist, by which term I mean any opponent of fatalism, whether motivated by reason of libertarian belief in free will or a compatibilist urge to come to some kind of accommodation of human freedom with causal determinism, or divine foreknowledge? Fatalists and anti-fatalists may both agree about the necessity of the inviolable past. The past is ontologically (if not always epistemically) closed. The anti-fatalist is however absolutely opposed to the notion of fatalism about the future.

---

24 For example, Michael Dummett has recently bothered to respond to this vulgar form of the fatalist argument (1986: 139-40), but if this is all the fatalist's position amounted to, then we might well wonder what all the fuss has been about down the centuries. He says of the fatalist that he 'is trying to persuade us that it is never reasonable to do anything in order to bring about, or prevent, a subsequent event' (1986: 144).
Anti-fatalists view the past and the future asymmetrically, and this goes a long way towards explaining why anti-fatalists often find theories of time which deny the ultimate ontological significance of any such asymmetry uncongenial (which is of course not to say that only those who reject such theories can be anti-fatalists).

What I shall demonstrate in the following chapters is that the answer to the truth about fatalism lies buried in the nature of the future and its relationship to the past. Consequently I shall argue the case for a particular type of fatalism, which I will call **ontological fatalism**. This type of fatalism sees the unavoidability inherent in all events as arising out of certain facts about the actual constitution of the world, rather than—as is more typical—from any idea of antecedently true propositions about the future.

I have alluded above to various views of time. Since these will prove to be crucial to the development of this argument, I propose to outline the two main contending theories in the next chapter. Following that, I will address the issue of their relationship to fatalism, both as I, and as others, understand it.
CHAPTER 2
THEORIES OF TIME

For time is like the fashionable host,
That slightly shakes his parting guest by the hand,
And with his arms outstretched, as he would fly,
Grasps in the comer. The welcome ever smiles,
And farewell goes out sighing. Let not virtue seek
Remuneration for the thing it was.

_Troilus and Cressida_ III, iii, 165-70

...the distinction between past, present and future is only an illusion, however persistent.

Albert Einstein, to Michele Besso's widow (1955)

...it is abundantly clear that neither the past nor the future exist...My problem was to discover the fundamental nature of time and what power it has.

St Augustine,
_Confessions_, Book XI, 23

2.1 Ordinary Ideas about Time

Time, as we ordinarily understand it, seems to involve at least the following two elements: _the experience of temporal passage_ and _the concepts of past, present and future_. These are related since the passage of time appears to involve events moving from the nonexistent future through the existing present to the nonexistent past. An event in the distant future comes eventually into existence in the present before coming to recede into the past. There are essential problems in making sense of the very notion of temporal passage, since passage seems to be a spatial concept—something passes by being at different places at different times. How then can we make sense of the notion of time itself passing?

One way in which philosophers have attempted to make sense of temporal passage is via the view that past, present and future all do exist or are real, but that they differ in their temporal properties. That is, events in these regions possess differing temporal characteristics: _viz_, **pastness**, **presentness**, and **futurity** are nonrelational properties which events possess. The passage of time thus consists in events changing with respect to these properties. Temporal passage or becoming thus comes to be seen as a type of qualitative change. But there is more to the ordinary conception of time than just passage. There is also the notion of succession. Our experience of time involves not only passage but also succession, in the sense that we experience events as occurring one after the other. Temporal relations
pertain to simultaneity and successiveness. But ultimately which is more fundamental—passage or succession?

2.2 Tensed and Tenseless Views of Time

Two contending theories have emerged: one theory sees our experience as depending ultimately on our experience of temporal passage, the other as depending on temporal relations. The former takes the transitory, nonrelational or monadic temporal properties of pastness, presentness and futurity (i.e., A-properties) as fundamental; the latter takes the relations of earlier than, later than, and simultaneous with as fundamental. These relations are polyadic properties, or B-relations.

Temporal becoming is obviously of central importance in the philosophy of time, but the debate over it can be formulated in various ways. One of these ways concerns the ontological status of events occupying different temporal locations. The issue then becomes: are events first future, then they become present, and finally become past (which they remain thereafter, only becoming more so as times goes on)? Or rather, it is wrong to speak of events 'coming into existence' when in reality they merely exist at their respective temporal locations. As Quentin Smith expresses it, '[i]f there is temporal becoming, then future events do not yet exist, present events exist, and past events no longer exist; but if there is no temporal becoming, then all events exist equally, regardless of whether they are located in 5,000 B.C., the twentieth century, or the twenty-fourth century.'¹

Philosophers who maintain that all successively ordered events enjoy the same ontological status are sometimes called 'detensers' and adhere to what is called the tenseless, or B-theory of time. For them, the nature or essence of time is captured completely by tenseless sentences like 'The battle of Jutland is earlier than the battle of the Somme'. Philosophers who maintain that temporal passage or becoming is real and fundamental are called 'tensers' and advance the tensed, or A-theory of time. They believe that we need tensed sentences such as 'World War III is yet to come' or 'Thermopylae happened long ago' or 'I'm just having dinner now' if we are to describe adequately the nature of reality. These theories have often been referred to as the A- and B-Theories since they arise (via Gale 1968) out of J.M.E. McTaggart's notion that time can be treated in one of two ways—as an A-series

¹ Oaklander & Smith (1994: 1).
(past, present, future) or as a B-series (earlier than, later than). They are so-called according to which series is held to be more basic or fundamental.

Everyone seems agreed that events possess B-relations; the question in dispute is whether they also possess A-properties. While a belief that events possess A-properties is probably sufficient to make one a tenser, it is not necessary. It is important to realise that some tensers (e.g., A. N. Prior) reject the idea that events possess these A-properties and prefer to characterise the tenser/detenser dispute in terms of whether some propositions change their truth-value over time or whether all propositions may be regarded as having permanent truth-values. What is common to all tensers though is the rejection of the idea that time is simply a system of successively ordered, equally real events, and that tenseless sentences are sufficient to describe this system.

Two Forms of Past/Future Asymmetry

Commonsense supports the view that there is a dual asymmetry between past and future. *Ontologically*, the past is real and the future is not. Once an event has happened it becomes part of the furniture of the world. It is genuinely real, ontologically finished; it is determinate, untouchable, unalterably a feature of existence. A future event, on the other hand, is not part of the world's furniture; its ontological status is sheer non-existence, non-being. Since the future fails to exist, past/present and present/future are differentially related: the present cannot be related to the future in anything like the way in which the past is related to the present because the future is just not there to be related to. In the second sense, there exists a *modal* asymmetry, the past is *fixed* whereas the future is *open*. This is closely connected to the previous asymmetry in that the past is fixed because it is actual. The future is open because in the present there exist potentially many different possible futures. It is this potency, residing in the present, that accounts for the modal uniqueness of the future. This potency is connected with *freedom*, in that there is nothing in the presently existing causes that determines the future, and with *time*, since the empty future becomes the actual present depending on what I (and others) freely choose to do. If the future is not (yet) real, then it is indeterminate and still to be created or formed, and this leaves room for the possibility that we agents might have a hand in creating or bringing it about.

The past then is fixed in a way in which the future, being open, is not. It is this intuition which lies at the heart of the belief that it is within my power to help determine how the future shall be, but that it is too late now to do the same with
respect to the past. Most of us are ordinarily inclined to believe that the past, as opposed to the future, is fixed, stable, unalterable, closed, and hence outside our control. This closedness of the past: accords with our commonsense intuitions, in that what has happened is taken to be in principle unpreventable, or as Ackrill has it, temporally necessary.\(^2\) As Aristotle observed parenthetically in the *Nicomachean Ethics* (1139b5-11):

Nothing that is past is an object of choice, e.g. no one chooses to have sacked Troy; for no one deliberates about the past, but about what is future and capable of being otherwise, while what is past is not capable of not having taken place; hence Agathon is right in saying

For this alone is lacking even to God,
To make undone things that have once been done.\(^3\)

It should be remembered that we usually only talk about the unalterability of the past when we want to contrast it with the future, just as Aristotle does here. But, strictly speaking, the relevant issue is not alterability, for the future is no more alterable than the past. What is relevant is the notion of determining the future, which is possible (even for a fatalist, I would contend), as opposed to determining the past, which most would believe is not in the absence of compelling reasons to subscribe to the possibility of backwards causation. It is with this distinction in mind then, that we should interpret the notion of the closedness of the past.

In sum then, there are two major contending theories of time: the A-theory and the B-theory. The A-theory holds that the temporal properties of *pastness*, *presentness* and *futurity* are mind-independent properties of events over and above such temporal relations as *earlier than*, *simultaneous with* and *later than*. Temporal passage is an objectively real feature of the world. The present is real in a way in which both the past and (particularly) the future is not. The B-theory maintains that time consists solely of these relations, and is in fact nothing more than these relations. The B-theorist denies an objective lapse or passage of time, and insists that all parts of time are equally real, existent or actual. The B-Theorist denies the A-Theorist's contention that there is an ontological difference between the past and the future which results in the following logical asymmetry between a past that is fixed, determinate, closed and a future that is open, a realm of genuine possibilities.

\(^2\) Aristotle (1963: 133).

\(^3\) Aristotle (1941: 1024). To say that something is 'capable of being otherwise' is just to say that it is contingent.
This asymmetry is more fundamental than the epistemic asymmetry whereby we, as it happens, typically know more about the past than the future.

2.3 Differing Views of the Past and Future

It is however a little misleading to talk univocally of the A-theory. There are at least two distinguishable views that deserve that name. They comprise two of the three standard views about the ontological status of our three temporal regions; the third being the B-theoretical view. These views are readily distinguished through their ontological perspective on non-present reality. They may be listed under the following usefully descriptive, yet barbarous names:

(1) Presentism This is the A-theory in its purest form. Only those things that exist at the present moment can truly be said to exist. Things that exist in the past, and things that do not yet exist cannot properly be said to exist at all. St Augustine famously observed in his *Confessions* that neither times past nor times future have any being.4

(2) Accretionism This is a variation on the above. Present and past things partake of full and determinate reality in so far as the former are currently occurring and the latter have occurred. They are to be contrasted with the events that constitute (will constitute) the future (whatever these may turn out to be). Consequently, future things cannot be said to partake of any determinate reality at all and are thus mere possibilia. C.D. Broad was at one stage of his career the most eloquent exponent of this view. His theory:

...accepts the reality of the present and the past, but holds that the future is simply nothing at all. Nothing has happened to the present by becoming past except that fresh slices of existence have been added to the total history of the world. The past is thus as real as the present...The sum total of existence is always increasing, and it is this which gives the time-series a sense as well as an order.5

Views (1) and (2) at least agree about the status of the future, and it is for this reason that they are best viewed as comprising two factions under the A-theory rubric. While the notion of an ever-growing reality is strange, it is easy to see the motivation behind the accretionist model: it speaks to our underlying feeling that the past and

---

4 'What is now clear and plain is, that neither things to come nor past are' (St Augustine 1968: 44).
5 Broad (1923: 66-7).
future are somehow ontologically different. Nevertheless, some philosophers have argued that it is incoherent. William Craig sees this move as being

... incompatible with the A-theory because then events which are to us present are future but nevertheless existent to persons who might exist in our past, a view which is characteristic of the B-theory. To such persons, the future is on an ontological par with their present and past up to the moment which is our present. Worse still, their real future grows as the moment of the present advances, which seems incoherent, since a person at tn cannot be said to have different ontologically real futures at different times.6

(3) Determinationism 7 This is the view that all events are equally real. As such it stands opposed to the claims embodied in (1) and (2). Contra (1), past and future things possess just as much reality as present things. Contra (2), future events as just as fully real and determinate as present and past events. This then is just the view that there is no essential ontological difference between the past and the future.8 I believe this is rightly taken to be one of, if not the, main defining doctrines of the B-theory. The present moment is, appearances notwithstanding, no more real than any other moment which either was, or will be, present in its turn. To deny full existence to entities which fail to exist now is no less objectionable than the denial of this reality would be to things which fail to exist here. 'Elsewhen' is just as real as 'elsewhere'. It is of course true that our temporal language or discourse is asymmetrical in its treatment of space and time. Spatial discourse is relational in ways that temporal discourse typically is not—though no doubt could be.

2.4 The New Theory of Time

The principal feature of recent philosophy of time has been the tussle over the so-called 'new B-theory'. This has spawned an extensive and ongoing debate. Oaklander and Smith edited a volume in 1994 which collected most of the relevant papers from the last decade and a half (many of them by themselves). The terrain for the staging of this battle is largely a linguistic and semantic one. Tensed language reflects temporal passage; tenseless language reflects temporal relations. It is now

7 Timothy Sprigge (1992: 2) is responsible for this term.
8 In a well-known passage, Hermann Weyl writes poetically: 'The objective world simply is, it does not happen. Only to the gaze of my consciousness, crawling upward along the life line of my body, does a section of this world come to life as a fleeting image in space which continuously changes in time.' (As quoted in McCall 1994: 20).
accepted by all sides that the B-theory cannot hope to provide a tenseless translation for each tensed sentence. B-theorists contend that they can do all that is needed, namely, provide tenseless truth conditions for tensed statements. It used to be thought that which theory of time was true turned on the question whether or not tensed discourse could be translated without remainder into tenseless discourse. That is, does our use of tensed language capture an aspect of time that tenseless language not only fails to capture but cannot capture?

I do not intend to enter the lists in this contest, nor will I defend one theory or the other (except where, *en passant*, I have occasion to indicate particular areas of difficulty for one approach or the other). Rather I will examine these theories only as they relate to the question of fatalism. For example, I will look in detail at whether or not the B-theory entails fatalism. My focus will also be on those features of the A- and B-Theories that have to do with their ontological claims concerning the future, since these are most relevant to the issue of the truth of fatalism. These features are especially the ontological and modal asymmetries referred to earlier.

2.5 Problems of the A- and B-Theories

In this section, I will allude briefly to some of the principal areas of difficulty facing each theory. My survey will be neither comprehensive nor deep. It is meant merely to provide minimal background for an examination of a difficulty that has long plagued the B-Theory—the imputation of fatalism. In doing this, I perforce skim lightly over some weighty questions, making detailed comment only on the relations between STR and theories of time since this relationship provides an illuminating context for the issues of the ontological status of the future and the question of fatalism.

*Problems for the B-theory*

I have already mentioned the difficulties experienced by B-theorists in capturing precisely the content of tensed discourse. Apart from this, it suffers from certain counter-intuitive consequences of views such as that the flow of time is merely subjective. At a very pre-theoretic level, everyone seems to know what is meant by the flow of time. The B-theorist who wants to deny both passage and the essential difference between different moments of time must therefore shoulder the burden of explaining precisely how it is that something so apparently real is actually not so. If passage is unreal, how can we account for the unreal appearance of passage? The physicist John Wheeler once observed that 'time is nature's way to keep everything
from happening at once'. The B-theory seems to suggest that in a sense everything has happened at once, hence the description often used to describe it—the 'block universe'. Why then am I at this moment in time rather than any other, or why doesn't everything happen at once if all events are equally real? There are of course responses available; for example, the B-theorist could reply, 'You might as well ask, why am I here rather than somewhere else?' or 'If all places are equally real, as everyone agrees, why doesn't everything happen here?' Nevertheless, such facile replies can mask the fact that there is real conceptual work to be done here.

In addition, there is something decidedly odd in the B-theory's desire to treat the past and the future on a par. Take, for example, the so-called 'Thank goodness that's over' problem. Even if the past and the future are both real, we have entirely different attitudes to them. You are to visit the dentist next week. This visit, on the B-theory will be just as real the day after it has occurred as it is now. Why then is that I dread it so now and feel such relief afterwards? Imagine that someone offered you the following choice: you can be in a situation where you have just suffered nine years of unbearable agony and will now enjoy one year of indescribable bliss, or be in a situation where you have just suffered nine years of bliss and now face one year of agony. The B-theorist should surely prefer the latter option, since all moments are equally real, and the good outweighs the bad by a ratio of 9:1. But who would join her? This asymmetrical attitude to the past and the future does not sit well with their alleged ontological parity.

Problems for the A-theory

For its part, the A-theory, while it accords well with many of our everyday intuitions about time, also faces real problems. Once you get beyond the vague feeling that time does flow, it becomes increasingly difficult to make conceptual sense of the notion: if it flows, how fast does it flow? One second per second? But rates of flow require time in order to make sense, so must there be so second-order level of time in terms of first-order time flows? As for the present, how extensive is it? Has real existence been reduced to a knife edge? Or, if the past exists, what is its exact ontological status? These are just a few of the difficulties that the A-theory has to grapple with.

---


For an excellent discussion of this problem, see the essays by D.H. Mellor, Murray MacBeath, and Brian J. Garrett in Part III of Oaklander and Smith (1994).
2.6 The Relativistic Woes of the A-theory

But one of the most pressing problems faced by the A-theory is its apparent lack of agreement with the Special and General Theories of Relativity. It is often maintained that both versions of the A-theory (i.e., (1) and (2)) are simply refuted by the Special Theory (STR). The relativisation of time central to STR would seem to put the A-theory at odds with our best physical theory. I will sketch just two instances of this conceptual dissonance arising out of STR's general relativisation of time. There is another reason for raising these issues: our discussion will bring up the issue of how the B-theoretic attempt to escape the inconvenience experienced by the A-theory in the face of STR perhaps carries the associated cost of fatalism. And it is precisely this implication which has traditionally been thought to be one of the principal difficulties of the B-theory.

(A) The phenomenon of time dilation

Mu-mesons produced in the laboratory decay after about $10^{-6}$ seconds (= 1 microsecond). These particles are also produced by cosmic rays striking atoms of air high in the earth's upper atmosphere. Since it is possible to estimate both how high in the atmosphere the collision occurred, and also approximately how fast the mu-meson is travelling when it reaches the particle detectors on the surface of the earth, it is possible to compute how long it took between the production of the mu-meson and its detection on earth, viz., about $10 \times 10^{-6}$ sec (= 10 microseconds). Given this disparity between their lifetimes and the length of the journey, how do any mu-mesons manage to make it to the surface of the earth before decaying? While there are many possible (and testable) explanations, the most favoured is that cosmic ray mu-mesons live longer than terrestrial ones because they are moving much faster, and according to STR, the faster something moves, the longer its lifetime.\footnote{This account of the mu-meson can be found in Geroch (1978: 59-62).} It would seem then that if time passes, it does so at differing rates for different objects—something which does not accord with our ordinary view (except metaphorically).

(B) The relativity of simultaneity

The doctrine of the relativity of simultaneity and the absence of a distinguished, or in any way privileged frame of reference, as found in STR, rules out the classical notion of an instantaneous, world-wide point of temporal becoming. In the Einsteinian context of a four-dimensional spacetime continuum, 'now' loses its objective meaning and only possesses significance relative to a given frame of
reference. Whereas in classical physics, simultaneity had been a two term, transitive relation, in STR this becomes a three term relation—an event is simultaneous with another at a given reference frame. Now it is true that the indexicality of 'now' is not subjective in itself, despite the fact that observers in different frames will assign different classes of events as being simultaneous with any given utterance of 'now'; and it is also true that this fixing on a frame does not thereby fix a present. However, there is a sense in which a subjective element enters in here nonetheless, since the temptation has always been to distribute the assignments of 'present' according to such simultaneity class determinations. Nor does the possibility that two observers might agree on the indexical application of 'now' to at least some shared events undermine this. It is true that another observer might shoot me, and we both utter 'now' at that place-time, and thereby refer to the same event, even though we are each in different frames of reference and therefore project different simultaneity classes. But our ordinary understanding of 'now' seems far stronger than this: we feel that if two observers can truthfully utter 'now' at the same spatio-temporal location, then they would thereby agree on the whole set of events that are simultaneous with respect to their frame of reference. And it is this assumption that STR calls into question.

This result has consequences for attributions of reality that may be made on the part of observers. The two observers in Figure 1.1 disagree over the reality of the distant event, say, the launching of a space fleet from Andromeda. For A, the fleet is already on its way towards Earth, but for B, the decision whether or not to send it has not even been taken yet. But both observers exist at the same spot (for all intents and purposes); their spatiotemporal locations are co-incident, and hence simultaneous. Therefore A is determinately real relative to B. And the event in question is determinately real relative to A. Surely this must also mean that the event is determinately real relative to B. How could it be otherwise? If we accept that reality is transitive, to deny the reality of the event which is real for A (who is in turn real for B) would be absurd. Yet it would appear this is precisely what we have to do if we allow some events to be real and others not. The following diagram and example may make this clearer:12

---

12 Penrose, whose diagram we follow, has obviously modified Putnam's (1967) example of a future space fight, which in turn is a relativistic take on an Aristotelian example with which we will be much concerned in Chapter 4.
Figure 2.1 (Adapted from Penrose 1989: 260): Two people walking past each other on Earth disagree as to whether the Andromedan space fleet had been launched at the moment they pass. 'Simultaneous space' refers to the set of events which each observer would regard as taking place at the same time.

But then, maybe reality just isn't transitive. (B-theorists who might think STR has vindicated their stance should consider this possible response.) If the simultaneity relation is not transitive (as the relativity of simultaneity seems to imply), then must not determinate reality also be non-transitive, and therefore relative to frames of reference? Reality thus becomes relativized to the same extent that simultaneity has been by STR. Sklar points to the reason for the A-theory's difficulty with adopting this line:

Now there is nothing inconsistent or otherwise formally objectionable about such a relativized notion of 'reality for', but it does seem to take the

---

13 Sklar (1977: 274-5 and cf. his 1985). There are other options that one might adopt. For example, one might decide to consider real only those events that are present here and now: this, in effect, reduces reality to a here-now point-instant. However this is solipsism with a vengeance is very unappealing. Or one might arbitrarily pick out a frame of reference and regard it as privileged. It will be taken to define what the present is, and hence what is real or not. This delivers 'a cosmic time with absolute simultaneity and a world-wide edge of becoming', as Craig (1991: 283) puts it. Candidates certainly exist, for example, the unique frame of reference whereby the universe is regarded as expanding isotropically. But this might seem a fragile hook on which to hang something as weighty as temporal becoming and the absolute distinction between past and future. In any case, there is some confusion in this approach, as fixing on a frame does not thereby fix on a present. (This has been pointed out to me by Graham Nerlich.)
metaphysical heart out of the old claim that the present had genuine reality and the past and future lacked it. For what counts as the present is only a matter of arbitrary choice, and so then is what is taken as real. At this point one can easily see why one would adopt, instead, the line standardly taken by proponents of relativistic spacetime and declare all events, past, present and future, equally real. For the distinction among them, being reduced to a mere conventional way of speaking, seems far too fragile to bear any real metaphysical weight.\textsuperscript{14}

A B-theorist will no doubt be attracted to this escape from the apparent absurdity that emerges in connection with Penrose's (and similar) examples. This is because he already subscribes to the full determinate reality of all events, past, present and future. The problem, however, with adopting this global temporal realism has always been that it carries with it the potential threat of fatalism. An unreal, open future is regarded as necessary protection against this threat. If we believe in the determinate reality of the past and the openness of the future upon which we expect to be able to act, we will want to take events in our backward light cone to be real, events in the forward light cone to be unreal, and then we may make some kind of decision (usually either all real or all unreal)\textsuperscript{15} concerning those events, like the space fleet departure, to which we are spacelike separated. Failing these strategies, how can we deal with the implication of fatalism? This is the topic of the next chapter.

\textsuperscript{14} Sklar (1985: 297).

\textsuperscript{15} Taking all events at spacelike separation to be unreal has the bizarre consequence that 'there will be events which are now such that they will be in my real past at some future time, but which will never have a present reality to me at all' (Sklar 1985: 300; cf. Putnam 1967: 246).
CHAPTER 3

THE FATALISTIC IMPLICATIONS OF THEORIES OF TIME

They say more people are laid low by time anxiety than by time itself. But only time is fatal.

Martin Amis

Our brief discussion of the relationship between the Theory of Relativity and both theories of time has raised the question of whether the B-theory is fatalistic. While the B-theory is regularly accused of fatalism by its critics, the A-theory is hardly ever so accused. In this chapter, I will examine Mark Bernstein's contention that, contrary to what we would normally expect given the A-theory's commitment to an open future, this view is in fact fatalistic. After rebutting this view, I then proceed to an examination of the more common claim against the B-theory. In the course of examining this claim, we will uncover important features of one typical form of fatalist argument, as well as suggest a viable reply to various criticisms of it.

3.1 Fatalism and the A-theory

Mark Bernstein has argued that the A-theory (or what he calls the 'Dynamic Theory of Time'), when taken together with what he describes as an eminently plausible notion, to wit, the idea that 'the existence of an object is required for that object to have any properties', entails fatalism about the future.¹ Now while not everyone may be willing to grant that this notion is anywhere near as plausible as Bernstein suggests (Meinongians, for example), I think many would; and in any case, as I shall show, his argument fails even if we grant him this. The temporal version of this actualist thesis runs thus:

\[(AT) \text{ Necessarily, if an object possesses a property at a time, then that object exists at that time.}\]

Instantiating \((AT)\) in line with the A-theory gives us:

\[(ATA) \text{ Necessarily, if an object moves at a certain time, then it exists at that time.}\]

---

¹ Bernstein (1989: 467). He notes that Alvin Plantinga has dubbed the thesis 'Serious Actualism'.

Given \((A_{TA})\), the argument proceeds as follows:

1. The A-theory holds that all events are always moving.
2. Therefore all events always exist. [From \((A_{TA})\) & (1)]
3. If all events always (at all times) exist, then they now (already) exist.
4. But if events already exist (i.e., exist prior to the occurrence of any of our putatively relevant actions) then, in accordance with fatalism about the past, nothing can be done to bring about their non-occurrence.
5. If the event in question is a future event, then the A-theory entails fatalism about the future.

As Bernstein is aware, the usual A-theoretic ploy—affirming the indeterminateness of future events—fails here, since 'indeterminateness does not provide us with a Future which is empty enough.'

But there is something strange here: if the A-theory is committed to the view that all events are always moving, and thus, via \((A_{TA})\), to the pre-existence of future events in the present, then it is possible to show directly that this theory implies a contradiction, and is thereby objectionable on grounds independent of any alleged commitment to fatalism. Take a particular future event \(E\)—my next birthday, for example. According to the A-theory, it possesses the intrinsic property of futurity; indeed it is one of the central claims of this view that events possess such properties. \(E\) possesses the property of futurity in virtue of its occurrence on that future date; yet it is claimed that \(E\) also exists now, in virtue of moving towards the present now, so presumably it also possesses the property of presentness. So \(E\) possesses both the properties of futurity and presentness.

The response can be made that events unproblematically possess the properties of futurity, presentness and pastness, albeit at different times. My next birthday is now future and will soon be present and thereafter past. This very question is often raised in discussions of McTaggart's argument for the unreality of time. In that context this appeal to second level predicates like 'future in the present' and 'future in the past' is usually held to be inadequate as it generates a regress of higher and higher level predicates, on each level of which the contradiction re-

---

emerges. Be that as it may, this ploy cannot be appealed to in this context. For what has been claimed here is not that E has the property of futurity at one time and the property of presentness at another, but rather that a future event now exists, in the sense of having these properties at the same time.

The claim that all events always exist or exist at all times is prima facie implausible. This becomes clearer if we look at the spatial analogue of this claim:

(Aș) Necessarily, if an object has a certain property $p$ at a certain place, then it exists at that place.

What is the analogue to the claim that all events always exist? All objects everywhere exist. But we would certainly not interpret this to be claiming that all objects exist at all places. If all objects exist everywhere then everything exists here! But of course what is true is that everything exists where it exists (i.e., at its own peculiar here). I don't see why the A-theorist, any less than the B-theorist in this case, need be committed to anything more than the claim that all events exist at the times at which they exist. But this strategy will only work so long as the A-theorist can reject claims like (1). For if all events (including future ones) are always moving, then they do exist now. That is, the A-theorist can deny that all events are always moving; and thereby deny that future events (pre-) exist in any harmful sense. The present is the point of temporal becoming, the point at which events occur and thereby come into being. She can hold that all (but only) existing events move since the future ones have not yet come into existence. Of course, the moment they do appear the A-theorist must say that they are moving but there is no problem with that (beyond whatever inherent problems there may be with the notion of time flowing). Bernstein's picture of A-theoretic reality is of a line of events shuffling on past the present. But anyone who wished to deny that the future exists could have these events coming into existence, as it were, on the run, or erupting into existence and thereby into motion as well. There are undoubtedly many problems associated with this view, but it is not obviously impossible. Alternatively some A-theorists prefer to view the now as doing the moving, with all events stationary; they would not subscribe to the thesis that all events are always moving. Whether or not the acknowledged existence (more or less indeterminate) of future events on this view reinstates fatalistic concerns is another question.
3.2 Fatalism and the B-theory

It is often claimed that the B-theory of time entails fatalism. Most philosophers would regard such an entailment as constituting sufficient reason for rejecting this theory. Many would agree with William Hasker that the B-theory is 'inherently fatalistic'.\(^4\) I believe that while most of the arguments that have been offered to establish this claim are defective, a case can be made for the claim that the B-theory entails fatalism; but I go on to suggest that this is not necessarily a reason for rejecting this theory of time.

The Standard Argument that the B-theory Entails Fatalism

Recall that the essential features of the B-theory are the following: (a) assertion of the fully determinate reality of all events, past, present and future; (b) denial of temporal passage, i.e., the notion that all events either acquire or lose the monadic, non-relational properties of pastness, presence and futurity. There are no privileged events and no privileged, objective 'now'. The claim we will consider is that fatalism follows directly from these B-theory features, particularly, the full reality of all events.

We are to imagine the existence of an ideal 'Book of Nature', conceived as containing a record of all true propositions. Since the B-theory implies that future events exist, it is appropriate to hold that this book will contain, inter alia, the set of propositions, corresponding to those future events, which are (now or already) true. An argument for fatalism can now be formulated thus:

\begin{align*}
\text{(6)} & \quad \text{There is a set of propositions about the future which are (now) true.} \\
\text{(7)} & \quad \text{If there is a set of propositions about the future which are (now) true, then the future is unavoidable.} \\
\text{(8)} & \quad \text{If the future is unavoidable, then no action is a free action.} \\
\text{(9)} & \quad \text{Therefore, no action is a free action.}
\end{align*}

I will examine three criticisms of this argument that have been advanced, not just because they pertain to the issue of its soundness, but because (a) in the course of responding to their misprisions we will be able to get a clearer picture of exactly what the fatalist is claiming, and (b) it will emerge that the necessary modifications

This approach denies that the future is real. For Peter Geach, 'Future-land is a region of fairytale. "The future" consists of certain actual trends and tendencies in the present that have not yet been fulfilled.' Some statements that are apparently about the future are in fact really only about the present. Geach would claim that something like 'S will do X at t2', when uttered at t1, really means the conditions obtaining at t1 are tending toward S's doing X at t2. In general, then, Geach's strategy is to re-interpret claims apparently about the future as claims that are really about the present. So, a statement like 'Lee will read Hamlet tomorrow' really means that conditions presently obtaining tend toward (or are presently tending toward) Lee's reading the play tomorrow. Obviously the world's tending at t1 toward S's doing A at t2 is wholly compatible with S's not doing A at t2. So 'Lee will read Hamlet tomorrow' can be true today even if, as it turns out, she reads nothing at all tomorrow:

Geach's idea is that was going to happen at the earlier time may not be going to happen at a later time, because of some action taken in the interim. This is the way we can change the future: we can and often do bring it about that it will not be the case that p, although before our action it was going to be the case that p; it was right to say, then, 'It is going to be the case that p'. Before the operation it was right to say 'Johnny is going to bleed to death from the injury': after the operation this was no longer the case. The proposition therefore that is admitted to be true is not then an authentically future proposition—which is as it should be, since the future isn't yet real—and Lee is thus free to do as she pleases, the true proposition notwithstanding.

Response. However this response is inadequate, as it is inherently implausible to suppose that statements apparently about the future are in fact really about the present. When I predict some future event I certainly think I am talking about the future and not just about some present tendencies. What else can we mean by the

---

5 The present objections are thus in a sense preparatory labour for the prime objective. In the next chapter I will examine the two principal criticisms traditionally directed against logical fatalism.

6 Geach (1977: 52-3).

7 Geach (1977: 50).
use of the future tense? Calvin Normore (1993) points to the difference between the progressive present and the simple future tense: 'will' is 'a natural way of simply predicting', whereas 'what is going to be the case' is something which 'changes as causes arrive and depart the scene'.

Geach's strategy is also vulnerable to the following objection. It seems entirely reasonable to say things like the following: 'The way things stand now, i.e., the way the world is now tending, it looks like Lee will read Hamlet tomorrow, but she won't in fact do so'. It is surely not inconsistent to say something like this. But on Geach's analysis such a statement will come out as something like 'The world is now tending towards Lee's reading Hamlet tomorrow but the world is now tending towards Lee's not reading Hamlet tomorrow', a conjunction which is obviously inconsistent.

Criticism 2

The crucial premise in the fatalist's argument, as it is presented here, is (7). It is ambiguous, and the ambiguity comes to the fore when we try to unpack the meaning of 'unavoidable'. What is it for the future to be unavoidable? Consider two times, t and t', where t' is later than t. Let t be the present moment, and t' is thus some future time. Consider also a set of propositions p₁ ... pₙ, such that any one of them, pᵢ, has the form 'state of affairs Aᵢ occurs at t". There are at least two possible ways to disambiguate (7). These two readings are:

(7a) It is necessarily true that if p₁ ... pₙ are all true at t, then it will be true at t' that A₁ ... Aₙ obtain.

(7b) If p₁ ... pₙ are all true at t, then it is necessarily true that it will be true at t' that A₁ ... Aₙ obtain.

According to J. Kvanvig, (7b) 'claims that there is only one possible future on the assumption that propositions about the future are true or false (now).'. He agrees that if (7b) is true, the argument cannot be resisted, but thinks it false on the basis of the following two 'intuitively obvious' claims:

8 Normore (1993: 85).
9 See Kvanvig (1986: 11-13). We might also care to speculate how Geach's analysis would deal with sheer predictions in the quantum-mechanical realm, where it may be wholly inappropriate to talk of the world's being such now that x is, as a result, going to happen?
1. There are true claims about the future;
2. There is more than one possible world even if there are true claims about the future.

Does (7b) deny the obvious truth that Nixon might have beaten Kennedy? Kvanvig maintains that the proposition *Nixon wins the presidential election of 1960* implies no contradiction: it is false but not contradictory.

Gregory Ganssle considers an argument essentially similar to the one I have sketched. He interprets (7) simply as (7b), his version having it that 'If a proposition corresponding to a future event is already true, there is only one possible future.' Like Kvanvig, he equates there being only one possible future with fatalism (as is clear from his third premise). The falsity of (7), thus taken, is what leads him to reject the argument: 'That a proposition corresponding to a future event is already true does not entail that there is only one possible future. What it does entail is that there is only one actual future.'\(^{11}\) Ganssle is quite right to reject this version of (7). The consequent does not follow from the antecedent.\(^{12}\) Indeed is it very likely false: at any given time, as Ganssle asserts, 'there are many futures which are possible' (p. 217). Therefore (7b) is unlikely to be true, and hence a poor candidate for providing the proper sense of (2).

What about (7a)? Does the argument employ it instead? Once again there is a problem. In order for the consequent of (7) to match the antecedent of (8), the necessity operator in (7a) will have to be distributed so that it reads:

(7c) If it is a necessary truth that \(p_1 \ldots p_n\) are true at \(t\), then it is a necessary truth that, at \(t', A_1 \ldots A_n\) obtain.

---

\(^{11}\) Ganssle (1995: 217)...

\(^{12}\) In passing Ganssle notes that if the truth of a proposition about the future was sufficient to entail fatalism then the A-theory would be just as fatalistic as the B-theory. He thinks this because he believes that 'even according to the A-theorist propositions corresponding to future events are already true even if these events are not yet real' (1995: 217). Not all A-theorists would agree however about the truth of such propositions: many would adopt (what was probably) the Aristotelian line that such propositions are not as yet either true or false. Indeed A-theorists keen to defeat fatalism often adopt precisely this stance, denying that the sort of propositions we are examining are true (though they may well become true later).
But then, in order to keep the argument valid, (6) has to become the claim that the true propositions about the future are not only true, but are necessarily true:

\[(6a) \quad \text{There is a set of propositions about the future } p_1 \ldots p_n \text{ that are (now, i.e., at } t) \text{ all necessarily true.}\]

But, it is argued, (6a) is merely a statement of the fatalist thesis, and the implausibility of this has been argued above in connection with (7b). Therefore, reading (7) as (7a) requires alterations in the argument to make it valid that result in the sheer assumption of the fatalist thesis in premise (6). On the basis of the above reasoning, Kvanvig claims that Taylor's argument either makes implausible claims or it begs the question.\(^\text{13}\)

**Response.** Those motivated to defend the B-theory from this implication can adopt either or both of two main strategies: (i) they can attempt to show that fatalism does not follow from what the B-theory is committed to, or (ii) they can (most would think uncontroversially) advance the claim that fatalism is simply false (or at least that the putative argument fails to establish it). Either way, B-theorists will imagine they have escaped an unpalatable outcome.

We should note that typically this battle is fought between two types of anti-fatalist: the A-theory anti-fatalist attacks the B-theorist via the imputation of fatalism, while the B-theory anti-fatalist tries to deflect the accusation. The fatalist watches from the sidelines. Whether or not Ganssle is a B-theorist, he does strive to defend the B-theory from the accusation of fatalism. In doing so, he makes two points: firstly, that (7) (i.e., our (7b)) is false, since it could be the case that a proposition corresponding to a future event be already true yet it not be the case that there is only one possible future; secondly, that there are in fact many possible futures. Let us assume he is right on both counts. The first claim serves to undo the critic's argument; the second goes further—it is tantamount, surely, to a denial of the truth of fatalism. The question of the truth or falsity of fatalism is not the main issue, but it is relevant: after all, the B-theory's entailment of fatalism would only be a sufficient reason for the rejection of that theory if fatalism were false. It is thus useful to consider the fatalist's view of this matter.

How then might a fatalist respond to Ganssle? He need not accuse Ganssle of just begging the question against him. Indeed he would be unwise to do so, since

\(^\text{13}\) Kvanvig (1986: 85-6).
adopting such a strategy might do no more than imply to the unwary that the fatalist has to deny the notion of many possible futures. But in fact the fatalist can admit that there always exist many possible futures: he would no more deny this than he would assert that the proposition in question expresses a necessary truth. His point is just that the event is unavoidable, and hence there is only one feasible or achievable future, namely, the actual one. Ganssle and Kvanvig do not explicitly define fatalism but they would probably accept something like the following: fatalism is the view that logic implies that there do not exist alternative possible futures (and hence no one can have free will). Despite its ubiquity, this characterization is too strongly formulated. The fatalist is really only committed to the view that whatever happens is necessary, in the sense of being unavoidable. I want to argue later that, surprisingly, there being only one actual future (which would appear to be uncontroversial) is sufficient to establish this view. But first we should point to a fundamental flaw in the argument we have been considering, and it is a flaw that infects both offered interpretations of (7).

This flaw involves the assumption that the necessity operator at work in the argument is one that should be read 'It is necessarily true that...'. For example, recall that it was claimed that the necessity operator in (7a) had to be distributed to the consequent in order for it to match the antecedent of (8). While this is true, it is not uncontroversial to cash out that operator in terms of necessary truth. Taylor (and other fatalists) would argue that 'unavoidable' is not adequately captured by the notion of necessary truth. In claiming that the states of affairs referred to in p1 ... pn (namely A1 ... An) unavoidably obtain at t', the fatalist does not claim that p1 ... pn are necessarily true. And one of the reasons why he is unlikely to subscribe to this is precisely because he acknowledges that there is more than one possible world. If the fatalist can countenance more than one possible world, then he is hardly going to allow this paraphrase of (7), given the common idea of some p being necessarily true iff it is true in all possible worlds. It is its truth in this (actual) world that makes it unavoidable on his view. That it may be false in other possible worlds does nothing to subvert this.

Criticism 3

There is yet another problem with the argument. The culprit this time is the notion of a proposition being now or already true. Prima facie it seems very strange that such a seemingly tensed notion of truth should figure so prominently in an argument purporting to implicate a tenseless theory of time with fatalism. It is often claimed that the fatalist relies on a notion of truth-at-a-time. It is an assumption that figures
prominently in many arguments concerning the B-theory and fatalism. For example, Cahn argues that the B-theory implies LEM, and that the truth of LEM means that all events are necessary (i.e., fatalism is true).\textsuperscript{14} His argument for this latter claim relies heavily on his premise (4): 'But in that case it must \textit{already} be true that a sea-fight will take place tomorrow, such that there is now no possibility that it might not, or else it must already be true that a sea-fight will not take place tomorrow, such that there is now no possibility that it might.' In this premise, the crucial claim is the one about a proposition being already true.

Certain arguments (whether to the effect that the B-theory entails fatalism, or simply for fatalism itself) make this commitment to an apparent notion of temporal truth even more explicit. For example, Bernstein:

\begin{enumerate}
\item The B-theory\textsuperscript{15} implies that it is forever true that event E will occur.
\item We cannot modify or undo the past (he equates this claim with \textit{past} fatalism).
\item Being able to make E not occur would entail being able to change the past.
\item Therefore, if the B-theory is true, we could not make it the case that E would not occur (i.e., \textit{future} fatalism is true).
\end{enumerate}

Assume, for example, that Bill climbs Mt Ainslie at \(t_{10}\). Thus it was true at \(t_0\) that Bill would climb Mt Ainslie at \(t_{10}\). Indeed, as David Widerker puts it, it was 'always true' or 'true from all eternity'\textsuperscript{16} that this event would occur. And if so, then it was not within Bill's power to refrain from so doing, for to suppose otherwise would be to suggest that Bill 'had it within his power to alter the past' or 'had power over the past'.\textsuperscript{17} Thus the fatalist who attributes to Bill at \(t_0\) the power to refrain is committed

\textsuperscript{14} Cahn (1967:27).

\textsuperscript{15} Bernstein actually eschews this designation, preferring 'the Non-Dynamic Theory of Time', but I mean by the B-theory essentially what he gives this designation to (1989: 461).

\textsuperscript{16} Widerker (1989: 97) and (1987: 229) respectively.

\textsuperscript{17} Widerker (1987:229) and (1989: 98) respectively. Interestingly, these two formulations are not identical, although Widerker seems to think they are, given that they figure in reiterations of the same argument. The former is a far stronger (and more controversial) claim than the latter. If backward causation is possible, one might consistently have power over the past in the sense of being able to bring some past event about, yet she might not be able to alter any past event. I might now be able to act retrocausally to help kill Kennedy but be not at all able to prevent his death.
to holding that it was not true at t₀ that Bill would climb Mt Ainslie at t₁₀, 'and this, the fatalist claims, is to have power over the past.'

Both Widerker and Bernstein are doing no more than unpacking what was implicit in premise (7) of our original argument (that if there exists a set of propositions about the future, then the future is unavoidable). This explication is useful but the essential point of concern is the role played by tensed ascriptions of truth, i.e., by the idea of a proposition being true at a time. Whether or not it takes the form of claiming that a given proposition is now, always, already, eternally or simply true prior to the event it describes, Cahn, Bernstein, Widerker and numerous others all construe the fatalist as relying on a temporal or tensed notion of truth. Bernstein elsewhere avers that 'the Fatalist [is] clearly committed to the notions of tensed truths and tensed facts.' These different formulations have their ultimate ground in the exegetical uncertainty surrounding Aristotle's initial portrayal of the fatalist's reasoning (as we shall see in the next chapter).

Be that as it may, this is an obvious point of attack for the B-theorist wishing to resist the implication of fatalism (and indeed for others wishing to undermine the fatalist). For they may simply claim that truth is essentially timeless, and while the state of affairs which makes a proposition true may be located in time, its truth is not. It is senseless to say things such as 'it is true at t₀ that Bill will climb Mt Ainslie at t₁₀'. Their reason is that it just does not make sense to speak of a proposition being true or false at a time, any more than it makes sense to speak of it being true or false at a given place. As A.J. Ayer puts it, '[i]t is not true at any special time, whether now or in the future but just true. To ask when it is true is to put an improper question.' And it is presumably no less improper than it would be to speak of the proposition that Hackett is a suburb of Canberra or that the speed of light is 186 000 miles per second as being true in Sydney (or indeed anywhere).

Response. But far from this being a problem for the fatalist, he can join forces here with the anti-fatalist B-theorist in his attack on temporal truth. The fatalist's argument cannot be derailed simply by pointing out that truth attaches timelessly to

---

20 Ayer (1963?: 237??).
21 It is important to realise that the anti-fatalist has to hold not just that propositions are eternally or atemporally (as opposed to temporally) true, but that 'eternal' here cannot mean 'omnitemporal', since
propositions, and thus that the fatalist's claim that it is true now that p is just ill-formed. If propositions cannot legitimately be true at particular times, then they cannot be true prior to the event they are about. But it was wrong of the anti-fatalist all along to imagine that the fatalist required this truth-prior-to-the-event. Why can't the fatalist just claim that by 'It is true now that Bill will climb Mt Ainslie tomorrow' he simply means that if anyone were to claim today Bill will climb Mt Ainslie tomorrow they would be saying something true? After all, to say that it is true in Sydney that Hackett is a suburb of Canberra is just to claim that were anyone to utter this proposition in Sydney (or anywhere) they would, at that place, be uttering a truth. And this does not entail that the proposition is true-at-a-place. Truth is timeless in an exactly analogous sense. (The question of course is whether this timeless sense might turn out to be no less conducive to the fatalist's purpose than the illicit notion of timely truth.) One might also argue that the expression of this truth (in the form of some given sentence) undoubtedly occurs in time, even if the truth itself is timeless, and perhaps this is sufficient for the fatalist.

Perhaps we are letting the fatalist off too lightly. Maybe our explication of the apparently tensed ascription has left out something important. Oaklander has argued:

The core of the fatalist argument is to move from a future tense proposition's being already true, or true prior to the time at which the event allegedly takes place, to the conclusion that the conditions that guarantee the truth of the future tense proposition already exist; that they exist before the allegedly "future" event takes place. Thus, if a future tense proposition for example, "A will occur", is already true, then at present there exists either (a) a "future" fact A will occur or (b) a set of causally sufficient conditions which "force" it to come about that A will occur. In either case, I cannot bring it about that A will occur, much less that it will not occur, and consequently my action is unfree.

---

22 Both Widerker (1987: 230) and McCall (1966: 272) endorse this explication as a legitimate response.

23 See Gaskin (1994: 88). I will argue below however than not even the expression in time of a timeless truth is required to secure fatalism.

24 Oaklander (1984: 200-1). It is not clear why Oaklander thinks that (a)'s or (b)'s obtaining—while either might prevent someone's precluding A's occurrence—would mean that 'I cannot bring it about that A will occur'. Might it not be that the only reason why it is a future fact that A will occur is
Now if any slide is allowed from pre-existing truths to pre-existing facts to pre-existing conditions for the truth of the future event, this is dangerous since it will lend credence to the idea of pre-existing events or states of affairs. Oaklander is quite right to object that any notion of pre-existing truths, facts or states of affairs is anathema on a B-theoretic view. A proposition $p$ may be true at time $t$ without there being any actual event or state of affairs at $t$ that makes $p$ true at that time.$^{25}$ This applies to all times, including the past, as Paul Horwich makes abundantly clear:

There need be nothing now (or here, for that matter)—no present occurrences—that make it true. What makes it true, if it is true, are certain events that took place [in the past] ... Similarly, even though it was true last week that I would now scratch my head, there was no event or state of affairs last week that made it true.$^{26}$

The B-theorist anti-fatalist will thus attempt to evade the accusation of fatalism by denying that future facts pre-exist in any meaningful sense in the present.

But is it the case that 'if it is already true today that I will bring it about tomorrow it already exists today, and I cannot bring about or prevent in the future what already exists today'? $^{27}$ We certainly don't allow this inference with respect to the past: it is surely true that what I say about the past now (or indeed at any time) is made true by what obtains in the past. From that fact that it is true today that Brutus killed Caesar, I would not claim that we should admit that there exist in the present facts about the past. We do not countenance past facts post-existing in the present, just as (quite rightly) Oaklander does not want to countenance future facts pre-existing in the present. Similarly, what I might say about Sydney is made true here in Canberra by what the facts are in Sydney, i.e., in virtue of how things are in Sydney. Analogously, what I say about the past is true now depending on how things were in the past, and what I say about the future is made true now by what the facts are in the future (i.e., by what they will be). Generally, a given proposition is true because the event it describes occurs at $t_n$ (or $s_n$, where $s = $ a spatial location or

---

25 See Oaklander (1982: 212): '...to affirm at instant t that "P" is true or "not-P" is true implies that a certain state of affairs exists (tenselessly) or does not exist (tenselessly), but it does not imply that the state of affairs in question either exists or does not exist at the time of affirmation.'

26 Horwich (1987: 30).

place). If the event occurs at $t_n$ it is true earlier that it will occur; if an event occurs at $s_n$ it is true elsewhere that it occurs there. So if anyone claims that it is already true today that $p$ then all that means is that their claim (which happens to have been made now) is true and it is made true in virtue of facts obtaining not now but in the future (where they properly belong on any B-theoretic account).

Oaklander believes that '[w]hen properly understood the Russellian view denies that propositions about the future correspond to facts that exist in time and it denies that existence is a tensed notion, or synonymously, that to be is to be present' and that fatalist arguments depend on these assumptions.28 I believe that the notion of being already true can be innocuously explicated and that the (acknowledged) truth about which facts make propositions true is fully consistent with the fatalist's position, since this position is founded solely upon the claim of the determinate factuality of the future (and the past). If the proposition that Bill will climb Mt Ainslie tomorrow is true in virtue of the (eternal or tenseless or timeless) fact that Bill climbs Mt Ainslie on 6 June 1997, then it is *that fact* that necessitates that he will so climb.29 It is *that* something is a fact that causes the problem, *whenever* that fact obtains. Or at least, such will be my claim.

*Three 'morals'*

There are, I believe, three conclusions we can draw from the inadequacy of these criticisms: first, that the reality of future events is critical (both to the fatalist argument itself, and hence to the claim that fatalism follows from the truth of the B-theory). Second, fatalism is not correctly construed as asserting that there is only one possible future, nor that all truths are necessary. Rather it asserts that there is only one actual future and that that is what is crucial. Third, fatalism relies not on a tensed theory of truth, but in reality on a timeless view, and comes over as a result of commitment to future facts, not pre-existing ones. It is interesting to note that these three claims are all consistent with the B-theory—indeed they almost define what it is to be a B-theorist. Let's see how the argument looks now, once it has taken on board these considerations.


29 A loose but suggestive analogy: I am upset to hear, while in Canberra, that my friend has been killed in Sydney. Do I feel any better to be told that, strictly speaking, this is not a fact that exists in Canberra, but in Sydney (or at no place at all)? Just as it is the fact of death that is emotionally upsetting—wherever it obtains—so it is the fact of the climb's occurrence that is the thorn in the B-theoretic anti-fatalist's side (in so far as the B-theory is committed to the existence of such facts).
3.3 The B-theory Does Entail Fatalism

So we will adopt the necessary modifications of the argument. The facticity to which we lately alluded is really nothing other than the B-theoretic commitment to a single actual future, comprising the set of all fully determinate future events. Accordingly we replace (7) with

(14) If a proposition corresponding to a future event is already true, there is only one actual future.

(14) is now uncontroversial (if we allow that future contingent propositions may be true, a question we may postpone until the following chapter). (8) must now change to

(15) If there is only one actual future, fatalism is true.

The argument that fatalism follows from a commitment to the truth of propositions about the future and the actual reality of the events these propositions concern is now quite straightforward:

(6) There is a set of propositions about the future which are (now) true.
(14) If a proposition corresponding to a future event is already true, there is only one actual future.
(15) If there is only one actual future, fatalism is true.
(16) Therefore, fatalism is true.³⁰

It is undoubtedly true that if there is only one possible future then fatalism is true, but this is far stronger than what is required. The fatalist needs merely the condition, which Ganssle himself advances, that only one future will be actual. Hence the actual existence of future events, as provided by the B-theory, is sufficient for the truth of fatalism.

Paradise on the cheap?

Many will, of course, be skeptical about getting fatalism so easily. In order to see how there being only one actual future might be sufficient to deliver fatalism, we

³⁰ I take this statement to be shorthand for the claim that no action is a free action, and hence (16) is essentially the same conclusion as (9) above.
need first to examine an additional argument that Ganssle discusses. He considers the idea that, even if the initial argument fails, 'there is something about the specific claims of the B-theory which renders it fatalistic.' This 'something' is precisely the actual existence of the future event, that is, what is asserted in the antecedent of (1). The crucial claim is that the future event is set and fixed in virtue of being a real feature of the future, and thus in a sense already exists. (Ganssle speaks very loosely of the event which exists tomorrow, i.e., the future event, existing today, but I think we can see what he means: he is emphasizing the fact that this fixed feature of the future is precisely this, even today.) Since the event is already 'set' in this manner an agent is not free to do otherwise.

The B-theorist anti-fatalist can respond to this argument by agreeing that Jeanie, to use Ganssle's example, will buy butter pecan ice cream tomorrow, but deny that she will be unable to buy chocolate instead. That is, she will buy butter pecan, but could have bought chocolate. The state of affairs that is the butter pecan purchase depends on her choice; it is not the case that the state of affairs constrains her choice:

The event [i.e., the purchase tomorrow] exists today because of the action Jeanie will perform tomorrow. If it were the case that she will perform a different action tomorrow, a different event would exist today. The event depends on Jeanie's purchase rather than the other way around.

This is a standard response to the problem of fatalism, and we will encounter it again, but is it adequate? It can only be so if it is true that future events are the result of libertarian free actions on the part of agents like Jeanie. If Jeanie is not free in this respect then Ganssle's point about the event's dependence on her purchase and her purchase's non-dependence upon the event is worthless. The claim has to be not only that if she were to act differently a different event would have obtained but also that she can do otherwise. A person is free with regard to some action \( A \) if and only if she has it within her power to perform \( A \) and has it within her power to refrain from performing \( A \). Let \( A \) be the action of Jeanie's choosing to buy butter pecan and \( \neg A \), her action of choosing to refrain from buying butter pecan (= choosing to buy chocolate instead). Let \( E \) be the (future) event that is this purchase of butter pecan and \( \neg E \) the (future) event that is her purchase of chocolate (and not butter pecan). Ganssle argues that because of \( A \), \( E \) obtains. \( A \) explains \( E \); \( E \) depends

---

on A in a way that A does not depend on E. According to the B-theorist either the conjunctive state of affairs (A & E) obtains or (¬A & ¬E) does. But if E obtains (and on the B-theory either E or ¬E does) then ¬A cannot; similarly, if ¬E is the existent event then A cannot obtain. The event that is the purchase and the action undertaken by Jeanie are both real and determinate features of the world. Neither is privileged in any way that the other is not. The fact that E obtains thus acts as a constraint on Jeanie’s power to act: she can act so as to bring about E (i.e., by doing A) but not so as bring about ¬E. For her to be free with respect to this action, she needs to have it in her power both to perform A and to perform ¬A. But she can only really have it within her power to perform one of these. Therefore she is not free and the future event is not the result of libertarian free action—which is just as the fatalist would have it. And this unfreedom is directly attributable to B-theoretic features of future events, viz., their determinate reality. It is also important to note that this situation is entirely consistent with A’s causing or bringing about E: fatalism is not about denying the efficacy of actions.

Why does the B-theory anti-fatalist imagine that an appeal to the dependence of future events on the prior actions of agents will help him avoid fatalism? There exists a residual feeling that at t1 it is up to the agent whether A or ¬A obtains, and that E or ¬E will occur at t2 as a result of which choice is made at t1. But this idea is infected with A-theoretic intuitions, specifically the notion that t1 comes before t2 in a more than merely relational sense and that A can be fixed prior to E becoming fixed. According to the B-theory, this perspective on things cannot confer a firmer reality on one event over another. The proper B-theory view is that neither A nor E is privileged in any way—if they obtain, they are both equally real, and neither can be properly thought of as somehow fixed prior to the other. No event or state of affairs is ‘harder’ than any other simply in virtue of occurring at an earlier time. It may well be that E is causally dependent upon A, and the proposition asserting its (i.e., E’s) existence is semantically dependent upon E, but this doesn’t affect the ontological truth that E is a real feature of the world-at-t2 at any time. Specifically,

34 Of course it may turn out occasionally that Jeanie determines on a choice of butter pecan but things conspire somehow so that she is actually prevented from making this purchase. But such anomalous instances of (A & ¬E)—or even (¬A & E)—cannot be appealed to by B-theory anti-fatalists since the very point they stress is that whether E or ¬E occurs depends precisely on the decision A or ¬A. We will make a presumption of efficacy and hence ignore such anomalies.

35 This last locution is not intended to suggest that E occurs at t2 at t1 (or any t)—just that time is irrelevant to the fact that E occurs at t2. I.e., were someone to have said at t1 (or at any t) ‘E occurs at t2’, then they would have spoken truly.
the fact that one event is causally dependent upon another does not, by itself, entail that an agent is free with respect to the cause.

Ganssle asserts that 'it is not whether a future event already exists or the proposition corresponding to such an event is already true that is relevant to fatalism. It is whether these are in any sense necessary.' 36 He is right about truth and also right about an event already existing (for a B-theorist, when it exists is irrelevant), but whether it exists or not is relevant. If it exists (in the future) it is an unavoidable feature of the world, which no prior action can dislodge, just as (and for precisely the same reason) no subsequent action can shift or modify a past feature of the world. The future event is necessary, not in the sense that a proposition concerning it is a necessary truth, but in the sense that it is unavoidable.

I conclude that the B-theory does entail fatalism, but that this provides no real cause for concern since this is not a fatal consequence. Indeed, in so far as this is a plausible theory of time, fatalism is correspondingly plausible. In the course of arguing this claim we have examined a number of the ill-founded strategies advanced by anti-fatalists. We are yet to encounter the two most venerable responses to fatalism. The first of these has it that fatalism can be avoided by denying, in keeping with our intuitions about the openness of the future, that all propositions about the future are true. The second response contends that the fatalist is guilty of a modal fallacy. To these I now turn, via an examination of the oldest argument for fatalism.

PART TWO

NAVAL LOGIC - PAST AND FUTURE

4.1 The Traditional Argument for Vaisheka

We examined a version of logical fiction in the previous chapter. The focus
of the argument for this logical fiction concept is best found in the
section Chapter 9 of Aristotle’s De Interpretatione (Chapter 9, 11. 37. 11. 38. 15)
formulation and return the traditional argument for Aristotle. It would
acknowledge the formulation of the Aristotle’s argument for his own purposes
and consequently agreed upon by the common correspondence of this property
property. It is

Aristotle argues from the principle of necessity of statements or
propositions about future events to the necessity of these events. If it is true
that something will happen, it cannot not happen and therefore must be
certainly happen.

if all propositions, including all propositions about present or future events, are
either true or false, then every event in the future is necessary.

An Aristotle

for speculations on future

events
CHAPTER 4
ARISTOTLE’S DISCUSSION OF FATALISM

Other prophecies were current, which were asserted to have been delivered hundreds of years previously. They had a most pernicious effect upon the mind of the vulgar, as they induced a belief in fatalism.

Charles Mackay
*Extraordinary Popular Delusions and the Madness of Crowds*

↢ Others apart sat on a Hill retir’d.
In thoughts more elevate, and reason’d high
Of Providence, Foreknowledge, Will, and Fate,
Fixt Fate, Free will, Foreknowledge absolute,
And found no end, in wand’ring mazes lost.


Fatalism is the thesis that whatever happens must happen. It claims that we lack the ability to alter the future, no less than the past. According to 'logical' or 'truth' fatalism, the reasons for this inability have to do with considerations relating to time, truth and logic. Specifically, this interpretation of fatalism maintains that from the mere truth of a proposition its necessity follows, in the sense that the state of affairs it refers to is necessarily unpreventable, and that from the mere falsity of a proposition its impossibility follows, in the sense that it is impossible to effect or bring about the relevant state of affairs.

4.1 The Traditional Argument for Fatalism

We examined a version of logical fatalism in the previous chapter. The *locus classicus* of the argument for this *logical* (or *truth*) fatalism is to be found in the famous Chapter 9 of Aristotle's *De Interpretatione* [hereafter DI 9]. In DI 9 he formulates and rebuts the traditional argument for fatalism, although neither his formulation of the fatalist's argument nor his own response to it, is unambiguously agreed upon by the countless commentators on this text down the centuries. It is usually thought that Aristotle argues from the present truth or falsity of statements or propositions about future events to the necessity of those events. If it is true that something will happen, it cannot not happen and therefore must necessarily happen; if *all* propositions, including all propositions about events in the future, are *now* either true or false, then every event in the future is necessary, and nothing will happen as a result of choice or chance. As Aristotle says, "[b]ut if it was always true
to say that it was so, or would be so, it could not not be so' (18b12). If we grant this move from 'was so' to 'could not not be so', then the conclusion follows by exchanging equivalent modal terms—if x cannot not be, then it necessarily will be. That this is a controversial move is at least clear on all accounts, whatever final characterisation may be offered. A detailed survey of the hermeneutical terrain of DI 9 can be found in Craig (1988) and Gaskin (1995). I will be concerned not so much with establishing the correctness of any particular interpretation as with a philosophical examination of the various responses to the argument.

What then is Aristotle up to in DI 9? In this text, he offers (and then rejects) two arguments for fatalism. An examination of these will help to sharpen our appreciation of the issues involved. But first we need some definitions:

**Principle of Bivalence [PB]** This is the semantic thesis that (it is necessary that) every meaningful, assertoric statement of a proposition is either true or false (where a false statement is merely the denial of a true one).³

**Law of the Excluded Middle [LEM]** This is the syntactic thesis that ‘p ∨ ~p’ holds for any substitution into the position occupied by ‘p’. (A semantic version of LEM would have it that the disjunction formed by a proposition and its denial (p ∨ ~p) is true, or that two contradictory propositions are not false together, since one must be true.)⁴

---

¹ All references to *De Interpretatione* Ch. 9 are to J.L. Ackrill's translation (cited as Aristotle 1963), but I include the standard Bekker line numbers as appropriate.

² See particularly Craig's Chapter One and Gaskin's Sections 2-14 (especially 2-4 & 8.)

³ I elide any subtleties involving reference to sentences, statements and propositions, and will speak interchangeably of statements or propositions as being true or false, since nothing crucial turns on it in this context. The debate about what are the proper bearers of truth values is thus one for another day.

⁴ See Gaskin (1995: 13 fn 2) for the distinction between syntactic and semantic versions of LEM, and for why the syntactic version is preferable. Kneale & Kneale (1986: 47) also define PB semantically but LEM syntactically. Cf. Lukasiewicz (1967a: 23), and also McCall (1966: 277) who argues that since LEM is a law of logic, LEM and PB 'should be considered as on quite different footings'.

Tarskian Truth Schema [T] 'p' is true iff p (where 'p' is a variable ranging over propositions and where the right hand p refers to facts or states of affairs, i.e., what makes propositions true).5

The 'Necessity of the Past' Principle [NP] This is the idea that the lapse of time imposes a kind of necessity on things [temporal necessity]. What is past is no longer revocable or alterable in virtue of being past.

The classic example of a future event is Aristotle's sea battle tomorrow. We will call the assertion that such a naval conflict will occur a 'future contingent singular proposition' (hereafter FCSP). It is 'contingent' in that it is about something that, one feels, either might or might not occur, and it is 'singular' in so far as it has to do with individual events rather than classes of lawlike phenomena. If we accept LEM, the following disjunction is true: either there will be a sea battle tomorrow or there will not be a sea battle tomorrow (i.e., p v ¬p). In the light of this, the two arguments involve the following two claims:

(A) The truth of either disjunct entails fatalism;
(B) The past truth of either disjunct entails fatalism.

There is some scholarly disagreement as to whether or not these two arguments are properly distinct. I follow Gaskin (1995) in seeing them as embodying essentially the same reasoning. But before looking at them in detail, I will outline the main responses that have historically been offered to them.

4.2 Competing Interpretations

1. The standard view 6

On this view Aristotle disables the fatalist by denying that FCSPs possess antecedent truth values. That is, Aristotle denies PB in respect of FCSPs while preserving

---

5 This is meant to capture the Aristotelian notion of truth as correspondence: 'To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true' (Metaph. 1011b26-7; Aristotle 1941: 749).

LEM. If PB is true; these two theses are equivalent; but if PB is false, LEM can still be true. The fatalist seeks to infer the necessity of a FCSP from its truth. Aristotle accepts the inference but avoids fatalism by denying that FCSPs are true or false (or perhaps, already true)—all the while allowing that a disjunction of contradictory FCSPs is true (indeed, necessarily true).

2. The non-standard view

On this view what Aristotle denies is not the truth of FCSPs, but their necessity. Aristotle rejects the inference from truth to necessity, but he restricts neither PB nor LEM. He distinguishes between the truth/falsity of a FCSP and its necessity, allowing the former but refusing the latter, and in particular the inference from the necessity of LEM to the several necessities of its disjuncts.

Non-standard interpreters typically see the fatalist's mistake as involving some manner of modal fallacy along the lines of illicitly moving from □(p v ¬p) to □p v □¬p. What Aristotle is said to deny with respect to future singular propositions, but affirm for past/present propositions, is □p v □¬p (where the modal operator stands for some sort of non-logical, or temporal, necessity). So, if p is a true

---

7 This interpretation originated with Anscombe (1967) and is argued by Linsky (1954), Strang (1960), Wolff (1960), Rescher (1963 & 1968), Williams (1980), Lowe (1980). Many commentators are hard to classify. E.g., Spellman (1980) criticises Ackrill's not unequivocal case for the standard view, but without making it clear where she stands; and McKim criticises both views (though he sees the standard view as restricting LEM rather than PB, 1971-2: 82) and holds that, for Aristotle, FCSPs are now decisably true or false (1971-2: 108).

8 Temporal necessity might less misleadingly be called irrevocability or unalterability. It applies to any event that has happened. A necessary event is thus one that has already occurred while a temporally contingent one is one that has not yet occurred, and presumably might not do so. The
past/present proposition, then p will be true with temporal necessity, i.e., though it is logically possible for p to be false, it is temporally impossible. However, for any true future proposition p, p will be true with temporal contingency, for it is still possible for p to be false. The non-standard view maintains that Aristotle held that FCSPs must be true or false, but contingently so.\(^9\)

### 4.3 The Structure of *De Interpretatione* 9

Aristotle's discussion of fatalism in *De Interpretatione*, Chapter 9 has a tripartite structure. Taking each part in turn, we can sketch an overview of what Aristotle has to say about fatalism, drawing particular attention to the competing exegetical claims of the standard and the non-standard interpretations.

**Part 1 - 18a28-34**

Aristotle argues that a certain thesis applies to statements about the present and past but not—or at least not in the same way—to FCSPs. That thesis, according to the standard view, is PB, and according to the non-standard view, is \((\Box p \lor \Box \neg p)\).

\(^9\) There is no need to deny or restrict PB because, as Donald Williams points out, Aristotle's argument 'that the admission of truth about the future entails...fatalism is an argument so swaggeringly invalid that the student can hardly believe that he meant it' (1951: 291). Aristotle's mistake was to equate 'It is necessary that either there is a sea fight tomorrow or there is not a sea fight tomorrow' with 'either it is necessary that there is a sea fight tomorrow or it is necessary that there is not a sea fight tomorrow', and to equate 'It is necessary that, if there is a sea fight tomorrow, there is a sea fight tomorrow', and to equate 'If there is a sea fight tomorrow, it is necessary that there is a sea fight tomorrow' with 'If there is a sea fight tomorrow, it is necessary that there is a sea fight tomorrow' (1951: 291-2). But both these transformations (from \(\Box(p \lor \neg p)\) to \(\Box p \lor \Box \neg p\) and from \(\Box(p \rightarrow p)\) to \(p \rightarrow \Box p\)) are invalid.

Note however that Aristotle (1963: 50-1) himself repeatedly says that the fatalistic consequences follow just if every affirmation and opposite negation is true or false (18a34, a37) or one true, one false (b27-9) without indicating explicitly at any time that he thinks the inference from premise to conclusion invalid.

\(^{10}\) My exposition of Aristotle largely follows Gaskin (1995), which is exhaustive and compelling (and which defends the standard view).
Part 2 - 18a34-19a22

Aristotle develops the fatalist's argument that, given the truth of a FCSP, its necessity follows. He rejects the fatalist's conclusion that everything comes about of necessity; that obliges him to reject either the fatalist's premises, or his argument. According to the standard view, Aristotle accepts the derivation of the necessity of a FCSP from its truth, but he rejects the premiss, that FCSPs have a truth value. Therefore he rejects PB for FCSPs. According to the non-standard view, he accepts the premiss, and here merely gives the fatalist's argument without rejecting it.

The fatalist's position is developed and rejected via a reductio argument, the grounds being: first, that it is incompatible with the facts of human deliberation and action in that 'what will be has an origin both in deliberation and in action' (19a8-9), and second, that in general it is incompatible with the existence in the world of contingency, by which he means that with 'things that are not always actual ... both possibilities are open, both being and not being, and, consequently, both coming to be and not coming to be' (19a9-11). It is interesting to observe that Aristotle has here identified three of the strongest intuitions motivating an inclination to reject fatalism: the feeling that it fails to do justice to our acknowledged abilities to deliberate and then act in accordance with our desires in order to affect the world, thereby instantiating some of these possibilities at the expense of others. (However, since this alleged contingency is precisely what is at issue from the fatalist's point of view, the appeal to this intuition is less than compelling.)

Part 3 - 19a23-b4

According to the non-standard view, it is here that Aristotle rejects the fatalist's argument by exposing in it the fallacy of operator shift. Aristotle accepts PB (sometimes conflated with LEM by some commentators). Aristotle is construed as rejecting the inference from □(p v ¬p) (i.e., an acceptable instance of LEM) to □p v □¬p. According to the standard view, Aristotle has already accepted the fatalist's argument in Part 2; in Part 3 therefore he rejects the fatalist's premiss that every statement is either true or false, all the while carefully seeking to preserve the truth of LEM (and indeed its necessity).

4.4 The Two Arguments

The two versions of the fatalist argument to be found in Part 2. The first seems to show why unrestricted PB leads to fatalism; the second how antecedent truth

---

11 Aristotle (1963: 52).
legitimises the ascription of necessity. The first embodies our original claim (A), namely, that the fatalist argument is based on the truth of a future-tense proposition; the second embodies claim (B), namely, that the fatalist relies on the necessity of a past-tense proposition. Respectively they proceed thus:

*Argument 1: 18a34-b9*

1. If something is white or not white, then it is true to say or deny that it is white (18b1-2).
2. But if someone asserts and another denies that Fp (i.e., it will be the case that something is white), then (at least and at most) one of these two people must, given PB, be right (18a35-9).
3. If one of them is right, then what he is right about - viz., the truth/falsity of 'Fp' - must be the case (18a39-b1).
4. Hence 'Fp' and '¬Fp' are in fact severally necessity/impossible (18a34-5, 18b4).

*Argument 2: 18b9-16*

5. If something is white now, then it was true to say earlier that it would be white.
6. If it was true to say earlier that it would be white, then it was true to say at all earlier times that it would be white.
7. If it was true to say at all earlier times that it would be white, then it was necessary that it would be white.

There are three features of this second argument which might seem to be relevant but which in fact are not, namely, the role of the (necessary) past, the actual utterance on any particular occasion of the prediction, and the move from 'true at any time' to 'true at all times'. Of these, the necessity of the past (NP) is most significant since it appears to warrant the inference from truth to necessity in (7). Let p = 'This is white'; Pp = 'it was the case that p'; Fp = 'it will be the case that p'; □p = 'it is unavoidably the case that p':

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(8)</td>
<td>p</td>
</tr>
<tr>
<td>(9)</td>
<td>PFp</td>
</tr>
<tr>
<td>(10)</td>
<td>□PFp</td>
</tr>
<tr>
<td>(11)</td>
<td>□[PFp → p]</td>
</tr>
<tr>
<td>(12)</td>
<td>□p</td>
</tr>
</tbody>
</table>

---

12 See Gaskin (1995:24, 26-32) where he argues this in detail.
Gaskin observes that while Aristotle accepts the necessity of the past (i.e. NP) and the modal principle used in this argument, and while it may well capture what he wants to say (even if he may wish to question the use of PB in (9)), NP is not crucial to the argument. Given that the first argument has no reference to the past (although at 18a35-6, it does make use of the future), why shift the argument into the past? The first argument then uses the future but not the past, and the second, the past but not the future. Gaskin concludes that the core of the argument must lie in what is common to both arguments:

The common element is that in both cases the predictor makes his prediction before the event to which it relates. The precise temporal relation of either prediction or predicted event to the present moment is irrelevant; what matters is their relation to each other. But Aristotle does achieve something important by incorporating the future into the first argument and the past into the second: he shows that the same argument works wherever the relevant events are located in the time-series. By presenting two versions of the fatalist's position, and varying the temporal perspective from one to the other, Aristotle could be said to be making just the point (on behalf of the fatalist) that it is the relative, not absolute, positions in the time-series of prediction and predicted event which are functional in the argument. The truth of the prediction necessitates the occurrence of the predicted event regardless of when, in relation to the absolute now, either the prediction or that event occurs.13

So Argument 1 deals with the assertion and denial of a FCSP, namely, a prediction that takes place now or later and refers to a still later state of affairs, and Argument 2 deals with the (prior) truth of a FCSP, namely, a prediction that has already occurred, and which relates to a state of affairs which obtains now. These differences about the temporal locations of the relevant predictions (i.e., present truths about the future and past truths about the present) are as Gaskin says, nothing more 'than rhetorically motivated variations on the same logical theme'.14 Both arguments then turn on the issue of antecedent truth. For Craig the 'issue is the semantic relation between a proposition and its corresponding state of affairs. It does not appear that the reasoning appeals to the necessity of the past in any essential.'15 Steven Cahn however maintains that there are two distinct arguments.16

16 Cahn (1967: 27-8).
Some commentators conflate the two, ascribing in the process a more significant role to NP than it deserves, for example Palle Yourgrau:

But of two such opposed predictions of the future, one (by excluded middle) must obtain, hence (by necessity of the past) be necessary. Hence, paradoxically, the future (including any aspect that we would normally take to be contingent and thus open to chance or deliberation) turns out to be as necessary as the past.17

However adequate or not the appeal to NP may be as expressive of Aristotle's intentions, it is on the face of it an odd argument. In what sense is the 'prediction' subject to NP? Truth is hardly an event or a state of affairs in the past. But are we meant then to accept that past statements of true propositions about the future warrant (more or less directly) the necessity of these future events via their indirect inclusion in the (necessary) past? In such cases it seems that what *would* be temporally necessary (in accordance with the necessity of the past) is simply that such a statement was made or such a proposition asserted, that is, the actual existence of the utterance. It is clear that this is not what Aristotle intends as he explicitly states at 18b36-19a1 that whether or not anyone actually makes these statements does not matter.18 But, were NP to be crucial to the argument, whether or not these statements were actually made would not only have to matter, it would in fact have to be almost the only thing that did matter. The only alternative would be if Aristotle had thought that it is the necessity of the propositional content of true, past propositions that leads to fatalism. It is difficult to see how this would work: while something like a particular utterance, being an event, appears to be the sort of thing that could be subject to the necessity of the past, the propositional content of the proposition itself is an unlikely candidate for subjection to that principle. 'It is fixed and unalterable that Smith asserted p' is quite different from 'The state of affairs spoken of by p is necessary'.

The fact that it is semantic necessity rather than NP that seems to do the relevant work for Aristotle is one of the reasons for preferring the standard interpretation. Gaskin argues very convincingly that we can express Aristotle's thought thus: while he does not accept 'p → ¬p' as a theorem, he does accept 'Tp →

18 Aristotle (1963: 51-2): 'Nor, of course, does it make any difference whether any people made the contradictory statements or not. For clearly this is how the actual things are even if someone did not affirm it and another deny it'. 
The necessity expressed in this theorem is what makes the past temporally necessary. It is often thought that the fatalist is trying to transfer this temporal necessity to the future (hence that belief that Aristotle's fatalist relies on NP). But there is no transfer; rather the fatalist grounds his argument in sheer facticity.  

4.5 Basic Version of the Argument for Truth Fatalism

In the light then of the essential similarity of the two arguments, we can generate a working version of the fatalist argument which proceeds from the antecedent truth of a FCSP to the necessity of the occurrence of the events described by that proposition. In so far as a FCSP may be levied for any given event in the future, necessity will apply to all future events:

(13) For every time \( t \) and proposition \( p \), either \( p \) is true at \( t \) or else \( p \) is false at \( t \).

(14) If (1), then there is now a true [false] proposition which states that a sea battle will occur tomorrow.

(15) If there is now a true [false] proposition to the effect that a sea battle will occur tomorrow then a sea battle cannot not occur [cannot occur] tomorrow.

(16) If a sea battle cannot not occur [cannot occur] tomorrow, the battle is necessary [impossible].

(17) If the sea battle tomorrow is inevitable (i.e, either necessary or impossible), then whatever will happen in the future is inevitable.

(18) Whatever will happen in the future is inevitable.

Comments on the Basic Version of the argument

In the previous chapter, I exposed some of the misconceptions of the fatalist's reasoning and argued against the feasibility of a number of suggested escape routes. Before I examine the two principal attempts to defuse this argument, I will make some specific comments on its various elements.

---

19 Gaskin (1995: 83). See the whole of Section 8 for his detailed case.

20 Or as Gaskin puts it, 'what is temporally necessary is what is now unpreventable. Taylor's point is just that facts are now unpreventable, whether these are past, present or future facts' (1995: 93, fn. 39).
1. Premise (13) is just a version of the PB. The remaining premises (with the exception of (17), but see below) are drawn from (18b:9-16):

...if it is white now it was true to say earlier that it would be white; so that it was always true to say of anything that has happened that it would be so. But if it was always true to say that it was so, or would be so, it could not not be so, or not be going to be so. But if something cannot not happen it is impossible for it not to happen; and if it is impossible for something not to happen, it is necessary for it to happen. Everything that will be, therefore, happens necessarily.21

2. Premise (14) seems to depend only on the existence of a proposition concerning the occurrence (or non-occurrence) of a sea battle tomorrow. Nor, as Aristotle realised, do we need to assume that any individual actually expresses any such proposition. The point is just that someone could. As to the question of the temporal notion of truth at a time: all that (14) is intended to express is the idea that the truth is antecedent to the event described.

3. Premise (15) is based on the (apparently) very plausible notion that if it is true today that a certain event will happen tomorrow, then there is nothing that anyone can do now to prevent that thing's happening tomorrow. However it should be mentioned that David Lewis, Ned Markosian and Peter van Inwagen are just three recent philosophers (in a very long tradition) who believe that this superficially attractive premise is false.22 In that case, something should be said to suggest why it might be thought plausible. The idea is that necessarily reality is such as a true proposition says it is, and we have already encountered it in the discussion of the two arguments above. For example, if it is true that 'A sea battle will occur tomorrow', then necessarily (in the sense of unavoidably) a sea battle will occur tomorrow:

The semantic relationship between a true proposition and reality is such that the two must correspond, so that if the future singular proposition \( p \) is true then the described state of affairs \( S \) must eventuate—otherwise \( p \) is not after all true.23

21 Aristotle (1963: 50-1).
The key notion is that of the necessity of a true proposition's correspondence with reality.\textsuperscript{24} It is the necessity that characterizes the correspondence relationship between a proposition and reality as described by the proposition. When a proposition is true or false then necessarily the things described in the proposition must be so, must obtain. Aristotle is not inferring that because a proposition is true, it is necessarily true. Rather he is maintaining that necessarily if a proposition is true then reality is such as the proposition says it is. A proposition and its contradictory cannot both be true then since in that case opposites would then obtain in reality, in violation of the Law of Contradiction. It therefore follows that if every proposition is either true or false, then necessarily either every proposition or its contradictory is true. The fatalist's reasoning then depends wholly on logical considerations:

The relevant necessity is that characterising the semantic relationship between propositions and corresponding reality. Since the state of affairs\emph{ must} correspond to the true future contingent singular proposition, the possibility of the event's equally happening or not happening is illusory.\textsuperscript{25}

It is this notion of necessary correspondence between states of affairs and propositions that truly express them that the fatalist relies upon in this argument. That which is truly stated cannot be avoided: 'For what anyone has truly said could be the case cannot not happen; and of what happens it was always true to say that it would be the case' (19a4-5).\textsuperscript{26}

4. Premise (16) just says that what cannot be prevented will necessarily happen, that is to say, it is inevitable (or unavoidable), a point it seems difficult to dispute. It is but an explication of what we mean by 'inevitable'. So the consequent expresses just another way of saying that the event's occurrence or non-occurrence is inevitable.)

5. Premise (17) is simply meant to indicate that this point about future propositions is generalisable. There's nothing intrinsically special about this future event. The occurrence (or non-occurrence) of a sea battle tomorrow is a paradigmatic instance of an event most of us would usually regard as possible yet avoidable; if it turns out to be unavoidable, then all similar cases will share this unavoidability, and a general fatalism about future events will ensue. (This premise is unexpressed in Aristotle's text but is plainly intended.)

\textsuperscript{24} This is what D. Frede calls \textit{relative semantic necessity} (Frede 1985: 40). It is to be contrasted with the logical necessity of LEM, i.e., the necessity attaching to \((p \lor \neg p)\).

\textsuperscript{25} Craig (1988: 35).

\textsuperscript{26} Aristotle (1963: 52).
6. It should be noted that it is not just from the truth of \( p \) that the fatalistic conclusion follows: it follows just as readily from the falsity of the proposition. That is, it is not simply the truth of such propositions that has fatalistic consequences, rather it is the fact that they possess a bivalent truth value at all. As Craig observes, '[a]n antecedently false proposition entails fatalism as much as an antecedently true one.'

We can see why this is so if we reflect on the situation where it is false to say 'There will be a sea battle tomorrow' and it is false to say 'There will not be a sea battle tomorrow'. Given the necessity of the semantic relation, both these propositions would necessarily then fail to correspond with reality. But this is absurd. Hence, since it is impossible that both be false, one must be true. So if \( p \) is false, then \( \neg p \) must be true, viz., 'There will not be a sea battle tomorrow'. This non-battle is every bit an event in the future as any battle that does take place.

7. The two main interpretations of the Aristotelian version of the fatalist's argument are significant principally because they issue in two contending explanations of what is wrong with the argument, and hence in distinctive strategies for avoiding the conclusion. The first focusses on PB; the second, on the issue of whether the fatalist is guilty of a modal error.

### 4.6 Strategy (a): Restricting Bivalence

Since the argument is manifestly valid, to avoid the unpalatable conclusion, one must reject (at least one) of its premises. The first and most obvious candidate for rejection is (13). Since perhaps the time of Aristotle, it has found favour with anti-fatalists who in effect thereby turn the argument for fatalism into an argument against PB.. (I say 'perhaps' since, as we have seen, it is a moot point whether or not this was Aristotle's preferred response to the fatalist.) This tack is favoured by those who see Aristotle as accepting the fatalist's inference that if every prediction is earlier true or false then everything is of necessity. But if bivalence delivers necessity then, since many future events are not (yet) necessary (19a7-32), the corresponding predictions cannot be true or false. It is clear that this way of construing Aristotle's

---


28 Haack offers an argument of this form (1974: 74).
project endorses the standard line on DI9. Apart from this, rejecting PB is one way of making the Aristotelian point that the future is (sometimes) open. 29

The implications of restricting bivalence

To deny PB for some future propositions is just to say that there are some propositions about the future relative to t that are, at t, neither true nor false. 30 Propositions about the past and the present, and those expressing future necessities, are either determinately true or false now, but propositions about future contingencies (i.e., FCSPs) are not determinately true yet—in the sense that they are not true or false in advance of the event they describe. This is to say that the future is, at least in part, open. Upholding PB is the same as holding that the future is closed: i.e., every proposition about the future relative to t is, at t, either true or false. To reject premise PB then is to claim that the future is (at least) sometimes open.

But LEM has been retained, since while it is not yet true that there will be a sea battle tomorrow, it is true that either there will be a sea battle tomorrow or there will not be a sea battle tomorrow. LEM is a law of logic which can find a home in many-valued logics which violate PB. (Interestingly, Lukasiewicz who revived the distinction between PB and LEM, and who pioneered three-valued logic, does not necessarily provide the best formalisation for this strategy, since his system jettisons both PB and LEM.) McCall however has pointed out that a many-valued logic can be constructed that rejects PB while retaining LEM. 31 McCall finds this an attractive option for another reason: this is because LEM cannot be denied save on pain of absurdity since, by De Morgan's laws, \((p \lor \neg p) \equiv \neg(p \land \neg p)\). Hence he prefers to retain LEM, and yet avoid fatalism, but at the expense of PB. The disjunction can be true (in accordance with LEM), although neither disjunct is (as yet) true, or false.

29 Lukasiewicz (1967a), Ryle (1954), Thomason (1970) and Storrs McCall (1966) have all endorsed this strategy. Subsequently McCall changed his mind about PB, at least in regard to future contingents (see 1994: 43 fn. 33).

30 What is being proposed is not so much the outright rejection of PB, but rather the denial of its universal application, and specifically its restriction to propositions about the past and present. Note that restricting PB is not the same as the medieval interpretation's view that Aristotle adaptsp PB, since that view still regards PB as universally applicable albeit in different senses (see fn 7 above).

31 (McCall 1966: 277).
Quine, who referred to this as 'Aristotle's fantasy', and others have poured ridicule on this approach.32

Perhaps the manoeuvre may even be pointless. Gaskin has suggested, following Haack, that 'Aristotle's adherence to unrestricted LEM might seem sufficient, even in the context of restricted PB and restricted Conditional Proof, to yield necessitarianism.'33 Let Fp = 'It will be the case that p'; and (Fp v ¬Fp) be a substitution instance of LEM:

\[
\begin{array}{ll}
(19) & Fp v ¬Fp \\
(20) & Fp \quad \text{Assumption} \\
(21) & □Fp \\
(22) & □Fp v □¬Fp \\
(23) & ¬Fp \quad \text{Assumption} \\
(24) & □¬Fp \\
(25) & □Fp v □¬Fp \\
(26) & □Fp v □¬Fp \\
\end{array}
\]

LEM, 19a28 19a23-4 21, vel introduction 19a23-4 24, vel introduction 19, 20, 22, 23, 25, vel elimination

Gaskin observes that Aristotle cannot accept this conclusion. He must regard both disjuncts as false (i.e., ¬(□Fp v □¬Fp)), since □¬Fp & □¬¬Fp (= ◊¬Fp & ◊Fp) must be his way of expressing the contingency of the future.34 To avoid this consequence, an Aristotelian must question the inference rule of vel elimination.

Putting this worry to one side for a moment, what kind of semantics might characterise this strategy? When we say of a FCSP that it is neither true nor false, this can mean one of two things: either it has some other truth-value (e.g., 'indeterminate', 'possible', 'indifferent'), or lacks truth value altogether, i.e., it represents a truth-value gap. It is generally thought that the latter conception better expresses the Aristotelian situation. In that case, any system appropriate to formalise the rejection of PB and the retention of LEM would have to be non-truth-functional.35

---


34 Gaskin (1995: 87) directs us to 19a9-10, 19a34, and 19b3-4 (Aristotle 1963: 52-3) in support of this claim.

But there are further problems even here. Susan Haack has demonstrated that while PB and LEM are distinct principles, either of which might be true of a given system without the other, they are connected in that if LEM is a theorem of a system in which Tarski's truth schema T holds then PB must also hold for that system.\textsuperscript{36} That is to say, you can only affirm LEM while rejecting PB if you also reject T—which is a high price indeed:

\begin{align*}
(27) & \quad Tp = p \\
(28) & \quad p \lor \neg p \\
(29) & \quad p \quad \text{Assumption} \\
(30) & \quad Tp \quad 27, 29, \text{def. of } = \text{ and MPP} \\
(31) & \quad \neg p \quad \text{Assumption} \\
(32) & \quad T\neg p \quad 27, \neg p/p, 31, \text{def. of } = \text{ and MPP} \\
(33) & \quad Tp \lor T\neg p \quad 30, \lor \text{ introduction} \\
(34) & \quad Tp \lor T\neg p \quad 32, \lor \text{ introduction} \\
(35) & \quad Tp \lor T\neg p \quad 28, 29, 31, 33, 34, \lor \text{ elimination} \\
(36) & \quad Tp \lor Fp \quad 35, \text{def. of } F'
\end{align*}

(36), of course, just is PB. Should you choose to abandon LEM, Haack also suggests that T alone is sufficient for PB in certain circumstances.\textsuperscript{37} It looks like we might have to abandon T as well as PB. It is at this point that the method of \textit{supervaluations} can be applied to Aristotelian truth-value gaps in order to assuage our misgivings. Not only is van Fraassen's system of supervaluations appropriately

\textsuperscript{36} Indeed we could almost say that there seems to be a general tendency for PB and LEM to collapse into each other in the presence of T, as Kneale and Kneale (1986: 47) recognise: given Aristotle's definitions of true and false, PB and LEM are obviously equivalent for if 'It is true that P' is equivalent to 'P' and 'It is false that P' is equivalent to 'not-P', 'P or not-P' is plainly equivalent to 'It is true that P or it is false that P'.

Aristotle never explicitly defines truth and falsity, but he does give some clear indications: 'to say of what is that it is not or of what is not that it is, is false, while to say of what is that it is or of what is not that it is not, is true' (Metaph. 1011b26-7; 1941: 749). These remarks offer a view of truth as correspondence whereby for any proposition p, p is true iff p corresponds to reality and p is false iff p fails to correspond to reality. Reality itself is essentially non-contradictory, therefore contradictory propositions cannot both correspond to reality. Hence, assuming PB, LEM follows: '...so that he who says of anything that it is, or that it is not, will say either what is true or what is false...' (1011b28; 1941: 749).

non-truth-functional, but also neither does T (i.e., Tarski's truth schema) always hold, since it is false that for some $\alpha$, $\alpha \rightarrow T\alpha$.\(^{38}\) (In any case, the supervaluationist already calls into question certain classical inference rules, such as vel elimination.)

Haack suggests that, if logic needs to be modified along these lines,\(^{39}\) an appropriate formalisation uses van Fraassen's method.\(^{40}\) The key idea is as follows: a *supervaluation* assigns to a compound well-formed formula (wff), some component(s) of which are truth-valueless, that value which all classical valuations would assign it, if there is one, and otherwise no value. These semantics are not truth-functional, since when in each case both disjuncts are propositions representing truth-value gaps, they would assign 'true' to '$p \lor \neg p$' but no truth-value to '$p \lor q$'. This allows us to retain LEM, since it would be assigned 'true' by all classical valuations. At the same time, however, there is no need to give a value to *all* disjunctions of truth-valueless disjuncts: for although '$p \lor \neg p$' would be assigned 'true' by all classical valuations, '$p \lor q$' would be assigned 'true' by some valuations and 'false' by others, and therefore is assigned no value by a supervaluation. Haack feels that while supervaluations are an ingenious method of dealing with wffs that are truth-valueless, they are not really appropriate since supervaluations are only plausible if it is supposed that the wff must be either true or false, though we either don't or can't know which. But if the wff has no truth-value,

...why should the fact that certain compounds of such wffs would have a truth-value whichever truth-value the components had, be any reason for giving the compounds that value anyway? The principle upon which van Fraassen's semantics rests, so far from being specially appropriate to the accommodation of items which lack truth-value altogether, is plausible precisely on the assumption that the items in question have truth-values, though perhaps unknown truth-values.\(^{41}\)

---

39 She doesn't however think such modifications are necessary, because she endorses strategy (b) (see Section 4.7 below).
40 For the method of supervaluations, see van Fraassen (1966, 1968 & 1969). Thomason (1970) has applied it to the Aristotelian picture of time in which the past is fixed and determinate while the future is indeterminate but growing. This Aristotelian model has a tree-like structure branching into the future but with a single trunk representing the past. At any point there are many alternative future possibilities but only one possible past. This is obviously akin to the A-theory view. Yourgrau (1991: 83-94) offers a critique of Thomason's approach.
But if this method is anywhere appropriate, it should be here, since FCSPs, while lacking a truth-value, will eventually acquire truth-values as their gaps get filled.  

**Reasons for refusing to restrict PB**

What all of this means is that this strategy may be costlier than expected since we may have to give up not only PB but also T and certain inference rules, not to mention certain attractively symmetrical ways of treating the world. Focussing just on PB, we may well question the point of the whole exercise: are we not giving up quite a lot in order to solve a particular problem, however irksome. Michael Dummett has argued that 'this strategy is ... far too heavy for the task.' 'Surely,' he claims, 'the fatalist's argument was a sophistry uncoverable without wading into the deep waters of metaphysics or the foundations of logic.' 43 Besides, as David Widerker has observed, it is a purely formal response to the problem, and thus somewhat *ad hoc*, since we are given no indication as to why we should drop an intuitively appealing principle other than that assuming it leads to an undesirable conclusion. 44 We may be entitled to feel that this response concedes too much to an argument that is otherwise flawed. To these general misgivings we may add more specific ones.

1. **The inherent plausibility of truth-value links.** A powerful objection to the restriction of PB derives from the apparent plausibility of biconditionals asserting truth value links between propositions: Doesn't the fact that *p* is true now imply that it would earlier have been true? That is to say, if someone had predicted yesterday that I would read *King Lear* today, and I did in fact do precisely that, then don't we want to say that that person said something true? A person who yesterday predicted the winner of today's race has surely made a true statement. It is agreed by all that the proposition, asserted tomorrow, that Brunellus won the race is true; but the earlier prediction that Brunellus would win seems to express exactly the same propositional content as the later claim—only the tense has varied. 45 Furthermore, the same facts that serve to make past and present tense propositions true also serve to make the appropriate future tense propositions true. Why should we grant 'Lee will read *Hamlet* tomorrow', uttered on June 4, a different truth-value status from that of 'Lee read *Hamlet* yesterday', uttered on June 6, when they both make

---

43 Dummett (1986: 141).
45 Epistemological considerations are obviously irrelevant in distinguishing between the two claims.
precisely the same claim, albeit from differing temporal perspectives, about the facts of the matter, viz., a reading of *Hamlet* on June 5 by a particular person? (This is in line with our argument in the previous chapter.)

Aristotle is unmoved by this argument because of his tensed understanding of truth and its correspondence with facts. Following Palle Yourgrau, we can characterise Aristotle's view of truth thus:

**Aristotle's Principle [AP]**

(i) Truth comes from correspondence with the actual (i.e., present) facts. I.e., truth involves correspondence between propositions and existing (where 'existing' means being present or existing now) facts or states of affairs.

(ii) The future (or at least its contingent part) is not now a fact. (If existing facts or states of affairs are present, and the future is not present by definition, then the future is not a fact.)

This principle reflects Aristotle's (A-theoretic) conception of the open future, and it explains why he must reject the truth-value links. It follows from AP that someone who correctly predicted yesterday it would rain today, did not speak truly at that time! Thus, even if what someone says will happen, does indeed occur, what they said was not true. And indeed it looks like it can never be true. My 'future-tense' prediction will not 'come true', rather a present-tense version of it will be true.

---

46 Wright (1993: 177-8) also questions the advisability of denying the truth-value links in order to avoid fatalism.

47 Yourgrau (1991: 82, 94).

48 Aristotle (1963: 53) says that 'statements are true according to how the actual things are' (19a32).

49 If this claim is implausible, how much more so is the suggestion of A.N. Prior that FCSPs are bivalent, but that affirmative versions of them are 'simply false, no matter how things turn out later on' (1967: 129). Cf. also his (1962: 38-40). This view has it that future tense propositions whose causes do not exist in the present are just false (prior to the times they are about). But perverse consequences ensue: this view entails that a FCSP that describes a state of affairs that becomes actual is nevertheless false. Thus if someone asserts 'Lee will read *Hamlet* tomorrow', and she does, we have to say that the prediction was nevertheless false, and this seems even more absurd than saying that the prediction was not true (i.e., that it was not the case that the prediction was true).

Oaklander notes that 'Prior's gambit preserves the law of excluded middle at the cost of abandoning the law of contradiction' (1984: 220). He observes that if p is undetermined, then 'it will not be the case that p' and 'it will not be the case that not-p' are both true (since 'it will be the case that
The problem is however that this view of truth appears natural and compelling if one accepts a B-theory view of actual time as a linear ordering of actually existent events. But of course Aristotle would not accept this view. Does this mean that the truth-value link condition begs the question against Aristotle? I think it can be shown that the truth-value link approach is inherently more plausible than the AP with its objectionable counterintuitive consequences. It is then not so much the case that truth-value links presuppose the B-theory as it is that any adequate theory of time ought to embody the insight they represent.

As is often the case in such matters, a move to analogous spatial examples is helpful. What I here (in Canberra) say obtains in Sydney, is true here in Canberra (indeed, true simpliciter) if it does obtain in Sydney. We saw that AP claimed that truth came from correspondence with actual present facts and that the future was not now a fact. Imagine a spatial analogue for AP:

**Spatial Version of 'AP' [SAP]**

(i) Truth comes from correspondence with present spatial facts (i.e., here-facts).

(ii) Elsewhere is not here a fact (i.e., not a here-fact).

At first glance this spatial analogue is implausible: do we really want to reject the idea that if (the state of affairs described by) p obtains at any place, then it is elsewhere true to say that p obtains? In one (trivial but misleading) sense it is correct. Propositions about Sydney are made true by facts about Sydney (which are here-facts when one is in Sydney). But in another (more significant) sense this is quite wrong: if the asserter of a true proposition about Sydney is in Canberra at the time of assertion, then that proposition is presumably not made true by any here-facts (i.e., facts about Canberra), but rather by how things are in Sydney. Similarly, anything pertaining to the existence of Sydney (or anyone there) isn't a here-fact when one is in Canberra, and is a here-fact when one is in Sydney; but, more deeply, anything pertaining to the existence of Sydney (or anyone there) is a fact here in

p' and 'it will be the case that not-p' are both false on Prior's view). But if 'it will not be the case that not-p' is true then 'it will be the case that p' is true. So 'it will be the case that p' is true and 'it will be the case that p' is false (1984: 220).

50 Yourgrau does not seem to see this when he says: 'If it rains today, it now becomes true to say that yesterday's prediction became true. But it remains that yesterday, before today arrived (with its rain), the prediction, lacking any actual, existent factual correlate, was without a truth value' (1991: 94).
Canberra in virtue of how things are in Sydney. Analogously, some future event isn't a now-fact today, but it is when now = (say) tomorrow. A future fact is a fact simpliciter—but in virtue of how things are in the future.

2. Inconsistent approach to the reality of the future. An Aristotelian who avails himself of the strategy we have been examining might be accused of inconsistency in so far as he does not—as AP(2) suggests—regard all propositions about the future as lacking truth-values but excludes causally necessary propositions. Aristotle does not hold that the future is wholly unreal or empty. It is clear on the basis of what he says in DI 9 that he subscribed to what Paul Fitzgerald has called the 'Halfway Theory' whereby certain regions of the future are real and certain other regions are unreal. The future is thus only partly unreal—it is partly real, but only to the extent that it is causally determined by the present or the past, and it is unreal in so far as it is not causally determined. Future events are causally determined iff their occurrence is entailed by truths about the present or the past along with true statements of causal laws. As Fitzgerald puts it,

... logically contingent future-tense statements are true now if the future is causally determined to turn out as they predict; false, if it is causally determined that the future will not be what they predict, and neither true nor false if the future is not causally determined either way. (1968, p. 424)

It is of course into the last category that FCSPs fall: these predictions of future contingencies are as yet neither true nor false since there is as yet nothing real or existent that could serve to make them true or false. Examples of such future events that are causally determined by presently existent states of affairs would be (for Aristotle) the future movements of the heavenly bodies.51 Williams thinks it 'a peculiarity of the Aristotelian argument to allow some reality to the future' and speaks of a 'strangely selective credulity' which credits 'sporadic real connections, spun between some events and not others, while rejecting the prosy dimensional manifold of juxtaposed fact.' 52 In a similar vein, Fitzgerald charges the Halfway Theory with 'making an arbitrary and unjustified distinction in truth-value status' between statements of causal laws and statements of future contingencies.53 He

51 Cf. Aristotle, De generatione et corruptione, 2.11.(338a15-338b6; 1941: 530-1).
52 Williams (1951; 294).
suggests that it would be less arbitrary to withhold truth from all (at least logically contingent) propositions about the future.\(^{54}\)

But perhaps Aristotle is not saddled with this metaphysical arbitrariness. Williams thinks that the source of Aristotle's problem lies in the confusion of determinateness with determinedness; \(^{55}\) but it is rather his critics who are confused about determinateness here. Aristotle can quite handily reply that there is truth about some future events—namely, those where it is causally settled now whether or not they will occur—but this doesn't thereby render them real and determinate aspects of the landscape of the future. They are causally determined by facts obtaining in the present; they are not made true by future facts and hence involve no commitment to any future reality at all. Propositions concerning them correspond to present reality. As Ackrill has noted in connection with DI 9 generally, Aristotle holds to a 'rather crude realistic correspondence theory of truth.'\(^{56}\) The reason why he does not want to ascribe truth or falsity to FCSPs is after all that there is not yet anything in the facts for them to correspond or fail to correspond to. But for certain truths about the future this is not the case: present facts render it settled and determined that what they refer to will come to be. They are true in virtue of presently existing states of affairs. Does this mean they are not really 'future' propositions at all? No: although made true by presently existing states of affairs, which causally ensure the future events, they are still about the future, not the present.

3. Inconsistent approach to the reality of the past. But while Aristotle may evade this first accusation of parity failure, a more formidable instance waits in the wings, in the shape of the past.

Let us start by asking why it is that the Aristotelian wants to withhold truth (or falsity) from FCSPs. Presumably it is because the states of affairs they refer to do not exist yet. But what about past truths, which refer to states of affairs that no longer exist? If future tense propositions are neither true nor false, then neither are past tense ones.\(^{57}\) For the events corresponding to past tense propositions no longer exist just as those corresponding to future tense ones do not yet exist. The difference between 'not yet' and 'no longer' hardly seems weighty enough to warrant preferential treatment of past truths.

---

\(^{54}\) Fitzgerald (1985: 561).

\(^{55}\) Williams (1951: 292-3).

\(^{56}\) Aristotle (1963: 140).

\(^{57}\) Craig (1991: 59) also makes this point.
But surely what I say about the past is made true (now, or indeed at any time) by what obtains in the past. Aristotle certainly believes in the full determinateness or fixity of the past (indeed in its temporal necessity). It looks like he also has to admit that there exist in the present facts about the past. Otherwise, how can we have truths about the past, given AP(1). But this is precisely the sort of situation he denies with respect to future facts. This asymmetrical treatment should not of course surprise us (given Aristotle's generally A-theoretic conception of time), but it does seem odd that there should be no apparent problem for Aristotle in countenancing past facts post-existing in the present while he rejects the idea (in AP(2)) of future facts pre-existing in the present (which explains why FCSPs are not yet true). B-theory anti-fatalists also attempt to evade the accusation of fatalism by denying that future facts pre-exist in the present (and also that past facts post-exist in the present), but less inconsistently since, firstly, they countenance unrestricted PB, and secondly, they regard fully real future and past states of affairs as underwriting the truth of their respective propositions.58 R.D. Bradley puts it succinctly:

The truth of a proposition about the future implies, not that something is the case, but that something will be the case, just as the truth of a proposition about the past implies, not that something in the past exists now but that something in the past did exist—in the past.59

Drawing together these observations about temporal regions like the future and the past and other spatial regions, the view I endorse has it that what I say about Sydney is made true here in Canberra by what the facts are in Sydney; what I say about the past is true now depending on how things were in the past; and what I say about the future is made true now by what the facts are in the future (i.e., by what they will be in the future). Aristotle's views on truth and correspondence are too simple-minded. A proposition is true because the event it describes occurs at \( t_n \) (or \( s_n \), where \( s = \) spatial place). If the event occurs at \( t_n \) it is true earlier that it will occur. If an event occurs at \( s_n \) it is true elsewhere that it occurs there. To the extent that this view is plausible, any restrictions on PB will be correspondingly implausible.

4. *Special Theory of Relativity.* Finally, what about the considerations arising out of the Special Theory of relativity that we encountered in Chapter 2? As we have seen, there does not exist an unequivocal means of dividing past and future events up so

58 On this last point cf. Baylis (1936: 162-3).

that all observers agree about the members of each set. Hence there is no univocal way of dividing the sets of propositions to which PB does and does not apply. Since everyone will have their own unique division of the set of propositions into subsets where PB either holds or fails to, the lack of parity inherent this approach is multiplied indefinitely: everyone will make the cut differently. Remember our two people crossing in the street on earth and disagreeing over the question of whether or not the Andromedan space fleet has left yet. Let's imagine they are both Aristotelians attracted by the option of restricting PB such that they divide propositions up into two sets, viz., those that are already true and those that are not yet true or false. The bizarre result is that these two people, sharing precisely the same metaphysical and semantic views of the world, and existing contiguously (if momentarily) at the same place and time cannot agree as to the constituent members of each set! This strangeness is not an outright reason for rejecting this approach, but it should render it considerably less attractive.

4.7  Strategy (b): The Accusation of Modal Error

Even were we willing to accept all the counter-intuitive consequences of attacking Premise (13), there is some question whether it would successfully undermine the fatalist's reasoning. I will argue later that truth is not the salient issue. But first, let's look more closely at what is involved in the escape route chosen by those who are either unwilling, or unconvinced of the need, to give up PB.

All the proposed responses to the fatalist that we have examined, both here and in the last chapter, have assumed that if FCSPs are allowed to be true (or false) prior to the occurrence of the events to which they correspond, then fatalism follows. Our next escape route questions this assumption. The opponent of fatalism now switches the point of attack and, abandoning truth as a battlefield, accuses the fatalist of committing a modal error. This response dates back at least as far as Cicero. It allows that predictions can be true or false prior to the time of the occurrence of the events to which they occur, yet denies that this entails fatalism. If it is true that Lee will read *Hamlet* tomorrow, then it is illegitimate to infer that she cannot fail to read it tomorrow; it merely follows that she will not fail to read it.60 And we ought not to conflate failure to refrain with inability to refrain. The claim is that the fatalist argument sees

(37)  It is true that X will do A

60  Cicero first made this point in De Fato (1991: 17-20).
as entailing

(38) X cannot fail to do A

when all (37) does entail is

(39) X will do A.

The fallacy is that of moving from '□(p → q)' to 'p → □(q)', or in this instance from '□(p → p)' to 'p → □(p)'. Specifically, the move is from the truth of p (i.e., Tp) to its necessity or inevitability. Since fatalism thus follows from fallacious reasoning, and not from PB, there is no need to reject PB (at least on these grounds).

As we have seen, this is a reasonable interpretation of how Aristotle proceeds in describing the fatalist's case. In response, it is claimed that from the truth of p the most one is entitled to infer is that the sea battle will occur, not that it was necessary or inevitable. Similarly, the fact of a proposition's being true entails only that its denial is false, but not that the denial is impossible. Another way of putting this is to point out that (39) does not have necessitas consequentis (i.e., the necessity of the consequent of a conditional). The fact that it is true that X will do A, does not entail 'necessarily, X will do A'. We are only entitled to a necessitas consequentiae (i.e., the necessity of the whole conditional—'necessarily, if it is true that X will do A, then X will do A'). This necessity is not seen as any limitation on the agent's freedom, since it amounts merely to the trivial claim that one cannot make a contradiction true (viz., that it is true that X will do A and it is true that X will not do A). I will argue that the fact about the world expressed in (39) however is sufficient to make a case for X's inability to avoid doing A.

In her discussion of Aristotle's second version of the fatalist's argument, Gail Fine also accuses the fatalist of this error, suggesting that "Necessarily (if it was always true to say of anything that it would be white, then it was white)'", but that he then slides the operator to the apodosis of the conditional.61 Gaskin, for one, denies that this fatalist is what the fatalist is doing: in fact, Fine has illicitly placed the necessity operator before the conditional in the first place:

...the fatalist says no such thing. He argues that necessity is a
metaphysical consequence of antecedent truth; he does not commit the
fallacy of operator shift, because he has no initial operator to shift.
(Alternatively, if we interpret the fatalist's second argument as relying on
the necessity of the past, then although the fatalist certainly would have
to start with something like Fine's principle, there would be no fallacious
operator shift: the fatalist would be entitled to the necessity of the
apodosis, because he would have the necessity of the protasis.) 62

So the situation is as follows: it is not the case that the fatalist has illegitimately slid
from applying an operator with wide scope to applying one with narrow scope.
Rather he argues for the necessity of the apodosis on the basis of the truth of the
protasis. It is therefore wrong to construe his ploy as move or less surreptitiously
moving from '◊(p → p)' to 'p → ◊(p)'. The talk of modal fallacy serves only to mask
what is really happening: the fatalist is arguing for inability, not simply deriving it
illicitly from the failure to perform some action.

Richard Taylor responds to such arguments against the fatalist thus: we are
all fatalists about the past and do regard the fact that it is true that, say, a seafight
occurred yesterday as sufficient for establishing the inability of anyone today to do
anything about its occurrence yesterday. In the case of statements about the past we
allow an implication from the truth of such statements to their inevitability, so why
not allow a comparable implication vis-a-vis statements about the future? Where is
the independent argument to establish the ability of someone to do anything about
the occurrence the future event? Taylor sees the accusation of modal confusion as a
'distracting irrelevancy':

The fatalist argument has nothing to do with impossibility in those senses
familiar to logic. It has to do with unavoidability. It is, in other words,
concerned with human abilities. The fact that a statement is true does
not, to be sure, entail that it is necessary, nor that all false statements
express impossibilities. Nonetheless, no one is able to avoid what is truly
described, however contingently, in any statement nor to bring about
what is thus falsely described...This inability goes far beyond the obvious
impossibility of making something both true and false at once.63

62 Gaskin (1995: 33). He also observes: 'Given that operator shift is not to be diagnosed in the
fatalist's second argument, it is likely that it is a mistake to find it in the first argument' (1995:33).
The fatalist and the anti-fatalist disagree over what is or isn't in one's power to do. The latter affirms, while the former denies, that we as agents possess the power to perform acts which we do not in fact perform (i.e., non-actual acts).

**The Counterfactual Solution**

How can we this resolve this impasse? William Lane Craig asks what it means to assert that acts that one does not in fact perform are nevertheless within one's power to perform. This is his answer:

Very simply, assuming that I am causally free, it seems to me that to claim that some non-actual action is "within one's power" is to assert the truth of a counterfactual such as:

1. I could do some action \( a \), if I wanted to.

Suppose, furthermore, that it is true that

2. Were I to do \( a \), then some consequence \( c \) would obtain.

From the fact that \( c \) will not obtain, therefore, it would be fallacious to conclude that it is not within my power to perform \( a \). It follows from the fact that \( c \) will not obtain only that I do not perform \( a \), not that I could not perform \( a \). But if I could perform \( a \), then \( a \) is within my power. The factors relevant to whether I could perform \( a \) are the antecedent and contemporary causal conditions bearing on my decision to perform or refrain from \( a \). What is not relevant is the truth value of future contingent propositions pertinent to \( a \). The truth value of such propositions is a consequence, not a condition, of what I do with regard to \( a \) and so reflects in no wise upon what lies within my power.64

The most significant point here is that two claims are involved: firstly, a claim about what would have been the consequence of having done a particular action, and secondly, a claim that I have the ability to perform that action.

Out of this analysis arises what we can refer to as 'the Counterfactual Solution' to the problem of fatalism. The designation is Craig's, although he credits John Turk Saunders with the basic insight. The original context involved God's foreknowledge and theological fatalism, but the solution applies equally to logical fatalism—some would say more than equally, since while there is still disagreement as to whether it works in the theological fatalist context, most commentators in that arena seem to think it suffices to defuse logical fatalism. The key notion is that it is within X's power to do something such that if X had done it, then a different proposition would have been true (or, God would not have held the belief that he in fact did hold). The important point, both with respect to the issue of God's earlier

---

64 Craig (1991: 159-60).
belief in, and the prior truth of, a particular proposition, is that the relevant counterfactual in these cases is a backtracker. Craig summarises this approach:

Once this counterfactual analysis of the way in which non-actual events are 'within one's power' is grasped, the heart of fatalism is removed...[W]e can causally determine the course of future events. We cannot in a similar way causally determine the course of past events. But assuming that our actions are not causally determined, we have the ability to act in ways other than we in fact act and were we to act in these other ways, then states of reality conditional upon those acts would be different than they in fact are. The chronological precedence or subsequence of those states with regard to the action is irrelevant...So long as we can act such that, were we to act in that way, the past would have been different, no threat to human freedom can raise its head. 65

Both the theological and the logical fatalist need to realise that precisely those previously existing states of affairs to which fatalistic import is assigned are consequences of the very actions over which they are supposed to hold sway. As Nelson Pike, who initiated the whole modern debate on divine foreknowledge and theological fatalism, expresses it:

Sentences are true because the things described in those sentences occur or obtain; not vice versa. Stated otherwise, it is the occurrence of the events or the obtaining of the circumstances that make sentences describing those things true; not vice versa. Thus if "Jones will do A at t2" was true at t1, this is because Jones does A at t2; not vice versa. Jones's action at t2 is what makes a sentence predicting that action true. 66

Since the semantic relation between a proposition and its corresponding state of affairs is non-causal, the terms 'makes' and 'because' here indicate merely the 'direction of analysis' rather than any causal relation, as Pike puts it.

The supervenience conception of truth

The counterfactual response is not particularly new, and it regularly makes new appearances in new and original garb. In a recent book, Storrs McCall has made use of the notion of a 'supervenience conception of truth' to explain the error of the fatalist. He finds this strategy attractive because it would enable the retention of an indeterministic branching model of the future (in keeping with an Aristotelian view

66 Pike (1993: 145)
of time) along with a commitment to bivalence.\textsuperscript{67} This supervenience theory of truth holds that truth supervenes upon events so that the truth today of a FCSP about tomorrow depends not upon what propositions are true today but on what happens tomorrow, or just before; consequently, worries concerning fatalism or logical determinism are unfounded.\textsuperscript{68} If we recall our earlier conceit of an Ideal Book of Nature containing all true propositions, we may wonder whether it is the case that something happens because it is written up in such a book, or if it is the case that it is so written because it is going to happen. McCall develops a theory which emphasizes this asymmetric dependency of truth on events and the non-dependency of events on truth:

These propositions [i.e., propositions about the future], according to the principle of bivalence, have been true, or alternatively have been false, since the beginning of time, and will remain true (or false) for all future times. But the truth or falsehood of a future-tense proposition need impose no unwelcome limitations on what people can or can't do, or on what may or may not occur as a result of the chance selection of a single 'actual' branch on the branched model. We shall say that the truth of an empirical proposition \textit{supervenes upon} events in the sense of being wholly dependent upon them, while at the same time events in no way supervene upon truth. Thus the truth of the proposition that X is in Warsaw town square at noon next Friday depends upon what happens next Friday, which in turn presumably depends on what X decides to do, and in this way the sting of 'logical determinism' is drawn. If X's actions next Friday somehow supervened upon, or depended upon, the truth of the proposition 'X is in the town square at noon on Friday, 30 March' then logical determinism \textit{would} have a sting, but this is not so. According to what we have called the 'supervenience' theory truth supervenes on events, not the other way round. The set of true propositions in no way determines what the future is like. Instead, what the future is like determines the set of true propositions \ldots Truth supervenes upon events: upon past events in the case of past-tense propositions, and upon future events in the case of future-tense propositions. But nothing supervenes

\textsuperscript{67} McCall (1994: 14).

\textsuperscript{68} McCall believes that Aristotle's own conception of truth supports the 'supervenience' theory: As Aristotle says in DI 9: 'it is not because of the affirming or denying that it will be or not be the case' (18b38-9; 1963: 52), and McCall continues, 'to which we might add, completing Aristotle's thought, "but because it will be or not be the case that we can truly affirm or deny today." Put in contemporary terms, what Aristotle appears to be saying is that truth supervenes upon objects and events in the sense of depending upon them, but that objects and events do not supervene upon truth.' (1994: 39 fn. 30) Cf. \textit{Metaph.} 1051b7-9: 'It is not because we think truly that you are pale, that you are pale, but because you are pale we who say this have the truth' (1941: 833).
upon truth, or in any case no events do. Truth 'bakes no bread'. It simply floats above the world, influencing nothing.\textsuperscript{69}

Given that some proposition about the future is true, is it possible to act in such a way as to bring it about that it be false? Fatalists like Taylor think not. McCall disagrees:

...it is possible to make a true proposition about tomorrow false. Since what makes it true in the first place is what we do tomorrow, and since there are many different things that we can do tomorrow, we can make it false. We won't in fact, as it turns out, but we can. Hence the type of fatalism forced on us by recognizing that there exists a body of true propositions about the future is entirely harmless, and in no way limits or restricts human freedom.\textsuperscript{70}

McCall's point comes down to this: 'What is possible or not possible for us to do is independent of what we will do, and hence of the truth or falsehood of the corresponding propositions. We can, therefore, do things that we will not in fact do.' What are we to make of this claim? It is merely asserted by McCall, or at least it doesn't seem to follow from what is possible or not for us to do being independent of what we will or won't do (and hence of antecedent truth). This is a concern since, in the absence of argument for it, it seems question-begging. McCall might appeal to its inherent intuitiveness, but the attractiveness of this appeal can easily be overrated. There are either reasons for accepting the intuition or there aren't. If there are, why bother with claiming that it is intuitively plausible; and if there aren't, then isn't the resort to intuition misleading?

Craig also merely assumes freedom. It may not follow from failure that I am unable, but this does not establish ability. We have still not been given any independent argument for ability, merely some explanation designed to render more palatable the idea that the truth of the FCSP ought not to be thought inhibiting since, had we acted differently, the truth would correspondingly have been other than it was. Furthermore, Craig's judgments as to what is and is not relevant to the question of what I can perform would simply be discounted by the fatalist—it is precisely this that is at issue. But the core of this response lies in the semantic observations about

\textsuperscript{69} McCall (1994: 14-5).

\textsuperscript{70} McCall (1994: 15-6). The answer to the theological fatalist is similar: 'It's not that our future actions are determined by God's knowing what we are going to do. In fact, it's just the other way round: our future actions determine what is true, and therefore determine what God knows. Human freedom, consequently, is in no way limited or constrained by divine foreknowledge' (1994: 16).
truth and conditionship. Are these as harmful to the fatalist as critics like Craig and McCall imagine?

Response to the Counterfactual Solution

The fatalist can admit the semantic point. The proposition is true because of the state of affairs; the state of affairs does not occur because the proposition is true. (But who ever—despite the imputations of anti-fatalists 71—seriously took the latter to be the case?) Even if we agree that the truth value of FCSPs is a consequence of the relevant state of affairs, the state of affairs itself, of which the true proposition is but a marker, might still act as an ultimately constraining factor. By itself, Craig's semantic point is fairly worthless: the fatalist in response should merely re-direct his attention towards what was always the real target.

Thus, had I acted differently, then the really existing state of affairs would have been different; hence the state of affairs referred to by the FCSP would have been different, and so a different FCSP would have been true. The world would have been different, hence so would the set of true propositions. But what does this really prove? The fatalist can readily admit the supervenience conception of truth, yet deny that the threat of fatalism is thereby obviated. It is true that 'truth bakes no bread'. But the lesson to be drawn from this is that truth was never really the issue; the important things are the ontological facts. It is the state of affairs itself qua determinate feature of the (future) world that is the problem. We can admit that propositions are true because the things described in those propositions occur, not vice versa (i.e., truth bakes no bread, as McCall puts it); but it is the thing that occurs, the event that is the performance of A, that circumscribes X's ability to act otherwise, in virtue of its obstinate rootedness in the real world, a characteristic to which the determinationist is firmly committed. Further, we should not forget that the anti-fatalist requires the ability to do A and the ability to do A', not just either ability. Given the state of affairs, whatever it may turn out to be, there is only one real possibility—the other is merely counterfactual.

The key question is this: What is the source of the threat to the ability claim, the true proposition or the ontological feature of the world, i.e., the actually existing

71 For example, Ryle (1954: 22): 'Lots of things could have prevented Eclipse from winning the race; lots of other things could have made his lead a longer one. But one thing had no influence on the race at all, namely the fact that if anyone guessed that he would win, he guessed correctly.' I assume (perhaps rashly) that a prediction's being correct is the same as its being true.
state of affairs, that the true proposition simply points to? Talk about truth is a convenient shorthand, but this should not mislead us into imagining that truth per se is the crucial issue. But we are ineluctably misled if we fail to realise the role played by competing (A-theoretic versus determinationist) visions of the future. We can now see how the problem was never really truth, but actuality itself. It is the ontological states of affairs, not semantic considerations, which ultimately determine ability claims and render the agent's action ineluctable. The fatalist should admit that what we will do (which is expressed by the antecedent truth) does not directly cause or determine what is is possible for us to do, yet this is not to say that these two things are wholly independent. We act in the context of a certain inventory of the world, one that includes all the events that occur at any time. This inventory must be self-consistent. Given that future events exist, it has not been established, pace McCall, et al., that we can do things that we won't in fact do. The fatalist is untouched by the counterfactual solution, no less—we now realise—than he would have been by the earlier 'truth' criticisms.

4.8 Time and Possibility

This line of response clearly owes whatever force it possesses to the notion that the actual existence of future states of affairs is sufficient to deprive us of real freedom. With Aristotle, we began with the question, does truth about the future entail necessity? In short, Aristotle's response was to admit it does and deny truth, while the B-theorist, rejecting the Aristotelian notion of an open future that informed that response, accepts truth but denies that necessity follows. The A-theoretic perspective of the Aristotelian sees the world's history as a single line up to a certain time $\alpha$ (= the present moment, now) with multiple branching lines representing possible future histories of the world thereafter. Take our sea battle (SB): both SB and $\neg$SB would inhabit these Aristotelian branches at some point beyond $\alpha$. But does the fact that there are alternative identifiable future histories from time $\alpha$, including $\neg$SB, show that, at $\alpha$, $\neg p$, viz., 'SB will not occur') is still possible, on the hypothesis that it is in fact the case that SB will occur (and hence that p is true)?

It all depends on what is meant by 'possible' here. The Aristotelian notion of possibility has it that something is now possible if it can yield to my deliberation, to my control in the sense of choice of action. But, by hypothesis, $p$ is true (i.e., the SB will occur) and thus $\neg$SB is not now subject to my control and deliberation, if these

---

72 But not only the B-theorist: some who are attracted to the A-theory—like Craig—also fall in with the B-theorist here, at least over the issue of modal error, if not the denial of the open future.
notions are taken to imply the real possibility of going either way. Therefore $\neg$SB is now not possible. The Aristotelian needs it to be the case that the world can go equally in one of two ways, SB or $\neg$SB, and that we can (at least help to) bring either about. Now there is obviously a sense of possible in which, given the assumption that SB will in fact occur (i.e., $p$ is true), $\neg$SB is possible, but it is merely a counterfactual sense. It is not the full-blown Aristotelian sense and this is not affected by the fact there are identifiable possible branches which by our hypothesis will not obtain but which, were they to obtain, would render $\neg p$ true:

In a word, the future is not open (to my deliberation and control) if it is already true, or a fact, what will happen, tomorrow, just because there are possible histories, or possible worlds, in which the future would be different. These histories/branches or worlds are possible not in the sense that they are (now) subject to my deliberation and control, but only in the sense that though they are not actual, they might have been.73

We must take care not to conflate these two senses of possible. If we already have truths about what will happen in the future (which the B-theorist requires) then we can only have possible branches in the restricted (as opposed to the full Aristotelian) sense. I.e., we can only have $\diamond p \lor \diamond \neg p$, but not $\diamond p \land \diamond \neg p$. And this seems to accord with our intuitions about future contingencies. In wondering whether or not a sea fight will occur tomorrow I assume that it is possible that either a sea fight will occur tomorrow or it will not (that is, $\diamond (p \lor \neg p)$); and while $\diamond (p \lor \neg p)$ entails $\diamond p \lor \diamond \neg p$, it does not entail $\diamond p \land \diamond \neg p$. In order for the disjunction to be true just one of the disjuncts needs to be true. The fatalist can quite happily accept $\diamond p \lor \diamond \neg p$ since he holds that whichever it is that is true, that one is necessary and hence unavoidable. In that case one of this pair will indicate a necessary state of affairs and the other a merely possible one. The Aristotelian can of course insist that real possibilities exist, that it is equally possible today that the sea fight occur and that it not occur, but this can only be sheer assertion.74 It is not justified on the basis of an appeal to our intuition that $\diamond (p \lor \neg p)$. And why should we pay the higher price that asymmetry involves when there is a less expensive alternative available? If this is the case we lose the branch asymmetry about $\alpha$. This will not worry the B-theorist.

---

74 And one that relies on the idea that some things are not actual, i.e., that the future is not (wholly) real in the sense that both possibilities—things being and not being—are open at that time. Cf. 19a9-11; Aristotle (1963: 52).
4.9 The Reappearance of Fatalism

However the question will now arise as to whether or not the restricted notion of possibility that remains is sufficient to underwrite the B-theorist's claim that his position avoids fatalism. For the B-theorist, it seems that there is no specifically past temporal necessity: if there are many (possible) futures but only one past (as the Aristotelian maintains) then, while we cannot distinguish past truth from past necessity (this is essentially what the concept of temporal necessity means), we can still distinguish future truth from future necessity, even if we decide, contra Aristotle, that future truth talk is acceptable. By hypothesis, we are taking the future to be just as much a fact as the past, with the result that there are in principle no more alternative possible futures in the relevant unrestricted sense of possible, than there are alternative possible pasts. This does not mean of course that we cannot make sense of the idea of alternative possible futures—just that they are to be construed along strictly analogous lines to the ways we already talk about alternative possible pasts. This is how we commonly speak about past events, e.g., it is logically possible that World War I might not have started in 1914 although in fact it did, and this fact is now properly seen to be unavoidable. Our model becomes then a linear one, representing the facticity of the future no less than the past. The not-to-be-actualized branches collapse. If it is claimed that we ought to retain these branches in any case, it must equally be accepted that we ought to countenance alternative possible, but non-actual past histories as well. The future is no more (or no less) open than the past.

For the B-theorist, this is all acceptable enough. He has understood that on the Aristotelian view present truth is wrongly held to go with present existence, and that 'Once we build into the fatalist argument the claim that (i) truth involves correspondence between propositions and existing states of affairs and that (ii) existing states of affairs are present, then the fatalist argument becomes very forceful'. The B-theorist thinks the problem can be avoided by realising that it is not the case that $p$ (or $\neg p$) need have any external correlate that exists at time $t$ in order for $p$ (or $\neg p$) to be true (at time $t$, or at any time), just so long as it is correlated with a fact obtaining at some time. All that the assertion at time $t$ that $p$ is true or that $\neg p$ is true implies is that 'a certain state of affairs exists (tenselessly) or does not exist (tenselessly), but it does not imply that the state of affairs in question either exists or does not exist at the time of affirmation'. In this case, the only legitimate

---

76 Oaklander (1982: 212).
remaining sense of possible is the restricted one involving mere ways-the-world-could-be-but-in-fact-won’t-be. But given this, doesn't this make it difficult for him to distinguish his position from fatalism (in the sense of the unavoidability of determinate factuality). For if 'A SB will occur tomorrow' is true (today) in virtue of the eternal fact that a SB occurs tomorrow, then it is that fact which necessitates that a SB occur tomorrow. Yourgrau seems to realize that necessity has not been defeated so long as you remain committed to factuality:

If the past is now necessary because it is, all of it, a big, unalterable fact to which present truths about the past correspond, then if there are truths about (all of) the future, then the future, too, is a big, unalterable fact, and so it, too, is in this sense, necessary. I detect here no sign of the modal fallacy that some have attributed to Aristotle.77

In this case, it looks like the existence of the eternal facts or states of affairs that the B-theorist is willing to countenance implies that no one has free will—a fatalist conclusion. The problem is that Oaklander thinks this also implies that it is not within one's power to prevent or bring about the SB’s occurrence tomorrow. And he rejects this latter claim: 'it is my later decision that accounts for which of two contradictory propositions about the future is true' and thereby accounts for which eternal facts obtain. He claims that the existence of these facts 'is not incompatible with our choosing or not choosing to bring about certain events', but that, 'on the contrary, it is because we do choose or do not choose to bring about certain events that certain [eternal] facts exist'.78

But to say that the existence of an eternal fact ensures that the event in question is unavoidable is not the same as to say that it is not within one's power either to bring about or prevent the said event. It is because it is an eternal fact that it is unavoidable, not because it might be brought about by me. How it comes to obtain is irrelevant to (and therefore obviously compatible with) its being an eternal fact. Indeed 'my later decision' is no less a fact on the B-theory than the future event (i.e., the SB). It is as fixed and determinate a part of the future as the SB. So choosing and not choosing are not on a par. Choosing and not choosing are not both real possibilities: one is actual and the other merely possible in the restricted sense. One is fixed and determinately actual; the other a merely possible way that the world might go but won't. To hold to any stronger sense of (real) possibilities obtaining is to appeal (illicitly for the B-theorist) to A-theoretic metaphysics.

We can admit that it is within our power to bring certain events about. And it is because we do have it within our power to bring about certain later events that certain later facts exist. But this power is not the same as the power to bring it about equally that either SB or not-SB exists. Our power extends to (◊SB v ◊¬SB), but not (◊SB & ◊¬SB). Showing that I am able to bring about SB is insufficient to warrant the claim that both SB and ¬SB are possible. Imagine I am in a room with two doors: one is a real exit, the other a sham exit. The fact that I can, by choosing (say) door A, bring it about that I leave the room is insufficient to establish that it is within my power to bring it about that I exit the room by leaving via door A or via door B. It is clearly not within my power to leave via door B, a fact entirely unaffected by the truth of the claim that of course it is logically possible that I might have left via door B had the world been relevantly different to the way it is. It is the fact that the door is impassable that is the constraining factor, not its position. Analogously, it is that x is a fact that causes the problem, whenever that fact obtains. It only looks like fatalism follows from x being already now a fact: this makes it look more inescapable, yet it is no less inescapable as an eternal fact.

But if the B-theory approach doesn't serve to avoid fatalism, oughtn't this move us to reconsider the Aristotelian solution? Yet it is difficult to see how restricting bivalence could help: it is the existence of the state of affairs that is the problem. To be sure, Aristotle's view of the open future is not ontologically committed to a determinate future. But why do we feel attracted to this view at all? Only because of our modal intuitions, that is to say, our feeling that the future could be different. Yet, as we have seen, the mere fact that there are possible future histories of the world in which the world would be different is in itself insufficient to establish this ontological openness. We can readily agree that the world could be different, yet stick with the unavoidability of the actual future. On the other hand, it being logically possible for persons to act otherwise seems insufficient to class them as free. It is a proposition's truth in this (actual) world that makes it unavoidable, and that it may be false in other possible worlds does nothing to subvert this, nor ground our freedom.

79 If someone objects that this does not capture all that we mean by freedom, or things being otherwise than they were, then I reply that this is all we ought to mean by these things. After all, that's all we mean with respect to the past: the actual past is unavoidably thus, although we can talk about what might have been.
4.10 Fischer's Use of Hard Facts

John Martin Fischer has argued in a number of places against a compatibilist line, such as Craig advocates, vis-a-vis divine foreknowledge and human free will. Craig believes that the solution to such problems is to be had in a manner analogous to the counterfactual solution to logical fatalism. Therefore it may be useful to consider Fischer's criticism of compatibilism in order to see how it differs from our response and whether it could be exploited by the fatalist as an alternative line.

Consider this example of Fischer's. Suppose the following two states of affairs obtain: (i) Mary stands up at t2, and (ii) God believed at tl that Mary would stand up at t2. According to Fischer, a compatibilist like Craig is committed to two claims:

(40) Mary can refrain from standing up at t2; and

(41) If Mary were to refrain from standing at t2, then God would have believed at t1 that Mary would refrain from standing at t2.

Fischer claims that the key issue is the relationship between the 'can' claim expressed in (40) and backtracking conditionals like (41). The truth of the backtracker is insufficient to establish the truth of compatibilism: 'I grant that the relevant backtracking conditionals are true. But it clearly does not follow that one has it in one's power to perform the relevant actions.' Indeed, Fischer thinks that it is at least plausible to suppose that the truth of the backtracker precludes the truth of the 'can' claim. How so? In so far as the backtracker claims that it is a necessary condition of X's refraining from performing A that some temporally genuine (hard) feature of the past would have been different from what it actually was, it is at least plausible to suppose that the former's truth rules out the truth of the latter;

In general, if it is a necessary condition of my performing a certain action A that I do something which intuitively I cannot do, then I cannot do A. And so it is at least plausible to suppose that if it is a necessary condition of my doing something that I so act that some temporally genuine—hard—feature of the past would have been different from what it actually was, then I cannot do the thing in question.

This relies on the notion of hard facts. The distinction between hard and soft facts is broadly a distinction between facts that are genuinely about the past as opposed to those which are only verbally about the past. A hard fact about a time \( t_n \) ought to be consistent with the ending of time immediately after \( t_n \), and hence does not require the existence of any later times. Fischer argues that the relevant facts about God's beliefs 82 are 'plausibly taken to have some "hard" (temporally non-relational) feature, and thus it is equally plausible to take them to be fixed at later times'. 83 (According to him, 'hardness' and 'softness' relate to temporal relationality, and 'fixity' to what an agent can control or affect.) Can these facts be sufficiently hard so that changing them would be something we intuitively cannot do, and thus we would be justified in denying that we could do the thing that required that they be changed? The truth of the backtracker seems to imply power over the past; Fischer thinks such power is something we cannot have, so we can't do that which was spoken of in the antecedent of the (admittedly perhaps true) backtracking conditional. A potential weakness in Fischer's argument is that there is some doubt about the hardness of God's beliefs. Whether or not this is so, there is considerable doubt as to whether the antecedent truth of a proposition is legitimately designated as 'hard', partly for reasons we have already discussed in relation to Aristotle's version of the fatalist's argument. Whether or not Fischer is right here in the theological fatalism case where the putative hard fact is the proposition that \( \text{God knows that } p \), it is very doubtful he is right when the putative hard fact is something like the proposition that \( \text{it is true at } t_1 \text{ that } p \).

The advantage of my response is that it sidesteps the whole reliance on hardness, around which a huge and increasingly epicyclic debate has grown. 84 Even if we have to accept that the past fact is not sufficiently 'hard' to do the job for the fatalist, we can retain the fatalist conclusion because the past fact (viz., the truth of the proposition, or the fact of God's (true) belief) merely points to the ontological state of affairs existing in the future. And this is the unavoidable fact or feature of

82 As I have already noted, Fischer is of course discussing divine foreknowledge, but the applicability to the case of logical fatalism should be apparent. In that case, the hard (or hard-type soft) facts would be the past assertion of some antecedently true FCSP.


84 For some of the flavour of this debate, see Craig (1991: Chapter 9); Fischer (1995: Chapter 6); Hasker (1989); and in particular the exchange between Craig (1989) and Fischer (1991 & 1992). For example, when Fischer starts to talk about "hard-type soft facts": soft facts with hard kernel properties' (1995: 118) one begins to feel that a once potentially useful distinction has got a little out of hand.
the world upon which fatalism builds its case. It is not the supervening truth, but the state of affairs upon which it supervenes that we have to worry about if we would resist fatalism. In effect, the real 'hard fact' is the future fact that corresponds to the state of affairs which, according to the B-theory, is ever real, and which makes true a proposition asserting that this state of affairs obtains (whenever it be asserted, or even if it never is).

4.11 Lewis on Fatalism

In the course of a classic discussion of the paradoxes of time travel (which we will examine more fully in a later chapter), David Lewis makes some interesting observations on the relationship of fatalists to the facts to which they appeal. It will help to sum up the lessons of this discussion, if we look at what he says there. He sees fatalists as people who argue that we can do less than we think. They 'take facts we count as irrelevant in saying what someone can do, disguise them somehow as facts of a different sort that we count as relevant, and thereby argue that we can do less than we think—indeed, that there is nothing at all that we don't do but can'.

I am not going to vote Republican next fall. The fatalist argues that, strange to say, I not only won't but can't; for my voting Republican is not compossible with the fact that it was true already in the year 1548 that I was not going to vote Republican 428 years later. My rejoinder is that this is a fact, sure enough; however, it is an irrelevant fact about the future masquerading as a relevant fact about the past, and so should be left out of account in saying what, in any ordinary sense, I can do.

I want to make two comments on this: first, Lewis's account of the fatalist's reason for arguing this way is misleading. True, the fatalist (and others) will want to say that it was true x years ago that I was going to vote Republican. But it is not because of this true proposition, uttered (or possibly uttered) in the past, that I am now unable to do otherwise. It is because of the state of affairs obtaining in the future that I am so unable. The true proposition that either was or could have been expressed years ago—Aristotle recognized that no-one need actually have uttered it—is true in virtue of the future state of affairs (under any plausible semantic account of its truth conditions). The fact about the future that is this state of affairs is therefore not masquerading as some past fact. It is a future fact, pure and simple.

---

85 Lewis (1986: 78).
86 Lewis (1986: 78).
Secondly, it is also not right to call it an irrelevant fact. It is relevant because it is what I will do. It is this fact that is (part of) the actual future. It is of course fallacious to conclude simply from the fact that \( e \) does not occur that it is not within my (or anyone's) power to perform some action \( a \) sufficient to bring about \( e \). It does follow that I do not in fact do \( a \). However, far too often the recognition of these truths is taken to be tantamount to an acceptance that it is therefore within my power to do \( a \). Now it is true that if I could perform \( a \), then \( a \) is certainly within my power. Craig says that '[t]he factors relevant to whether I could perform \( a \) are the antecedent and contemporary causal conditions bearing on my decision to perform or refrain from \( a \). What is not relevant is the truth value of future contingent propositions pertinent to \( a \). The truth value of such propositions is a consequence, not a condition, of what I do with regard to \( a \) and so reflects in no wise upon what lies within my power.'

But causal factors do not operate in a vacuum. Certain constraints apply. Among these are not, to be sure, true propositions. But there are, on any theory that upholds the notion of an actual future, actually obtaining (future) states of affairs. There is a consistency constraint that comes into play when theories like the B-theory are around. It is in fact the highest constraint on any such account. It is not the true propositions that constrain what agents may do; it is the facts to which they are semantically linked. An agent always performs an action within a set of circumstances, and it is just a mistake to imagine that these must always be circumstances obtaining at the time in question.

Imagine the following case: I believe I am to inherit a certain amount of money from a recently deceased uncle. With it I plan to build a gazebo in my garden. However, although my uncle always promised that I would get this money from his estate, in fact he died penniless, having gambled away all his money. (For the sake of the argument we can rule out other forms of income for me.) Now it seems plausible to say that when we are assessing my ability to construct the gazebo, we ought to take into account the fact that I will not be receiving any money from uncle's estate. That I may not realise this at the time is clearly irrelevant. My abilities surely ought not to be consequent upon my epistemic status at any time—although how or whether I choose to exercise them may be. It seems that we have to say that I cannot build the gazebo, not just because it is true for me here that there is no money in my uncle's account (although in some sense this is true), but rather because of the state of affairs obtaining in that account, namely, that it is empty.

---

Now of course we can admit the truth of all sorts of counterfactual statements to the effect that my uncle's account might not have been empty at his death, or that it might not have been the case that this was my only source of funds, and so on. Had these things been true, my abilities might well have been different. But they weren't, so I can't build the gazebo. In the circumstances (and we are always in some circumstances or other), I lack the power to build it. I contend that just as the state of my uncle's account on the other side of the world is relevant to the assessment of my ability to do a certain thing here, so too the state of affairs obtaining at certain places in the future is relevant to the assessment of my abilities to do certain things now.

Consequently, it is by no means certain that I could perform \( a \). And thus it has not been established that \( a \) is within my power. But this is a debate to which we will have to return in the next chapter, for it will arise in the context of the 'modern' classic argument for fatalism, namely, Richard Taylor's 'condition-fatalism'.\(^88\)

\(^{88}\) The name is Seddon's (1987: 110).
CHAPTER 5

RICHARD TAYLOR'S ARGUMENT FOR FATALISM

'Sir, (said he,) we know our will is free, and there's an end on't.'

'All theory is against the freedom of the will; all experience for it.'

'But, Sir, as to the doctrine of Necessity, no man believes it. If a man should give me arguments that I do not see, though I could not answer them, should I believe that I do not see?'

Samuel Johnson,
Boswell's Life of Johnson

Richard Taylor is by far the most profound contemporary fatalist, in what is admittedly a rather small field. He is responsible for the two most influential versions of the fatalist argument in modern times. The first version, following Aristotle, rehearses an argument based on the existence of a body of truth, or set of all true propositions. We have already examined in Chapters 2 and 3 the truth fatalism that relies on this type of argument. In his (1989), Taylor tells the story of a man called Osmo who discovers a book in which are written all the facts of his life. Osmo thereby learns, inter alia, the time and circumstances of his own death in a future plane crash. In a vain attempt to avoid the death he sees described in the book of his life, he goes berserk on board an aircraft flight, thus bringing about the very fate he was trying to escape. Taylor observes that, however hard some may have tried, no one has ever managed to render false a statement that was true.

The second type of argument that Taylor deploys, and the one that I will examine in this chapter, focuses on the idea of the relevance of past and future conditions to decisions about what agents can and cannot do. Accordingly, it is appropriate to refer to this version as an argument for condition fatalism.

Fatalism, as we have seen, is the doctrine that propositions like

(A) It is within an agent's power to do A, and it is also within an agent's power to do \( \neg A \)

are never true. Common sense and libertarians assert that many instances of (A) are true. Richard Taylor has argued that, on the contrary, such propositions can be shown to be false, even allowing for the fact that the defender of (A) does not mean
to assert that the agent can do both actions at once. Let us examine Taylor's reasoning.

5.1 The Argument for Condition Fatalism

Taylor's argument proceeds through two stages, each involving the explication of a situation or scenario: the first in which an agent is to perform an act of reading a newspaper headline about a seafight which may or may not have occurred yesterday, and a second in which the agent is an admiral who is to issue an order which will be sufficient either for bringing about or preventing a seafight tomorrow. In the first scenario, Taylor examines one instance of (A):

\[(AR) \text{ It is within an agent's power to do an act of reading } R, \text{ and it is also within an agent's power to do } \neg R.\]

In the second, he focuses on this instance:

\[(AO) \text{ It is within an agent's power to do the act of issuing an order } O, \text{ and it is also within an agent's power to do } \neg O.\]

Taylor will argue that \((AR)\) is false (which will surprise very few), but also that \((AO)\) is false (which is far more contentious, in so far as the denial of \((AO)\) constitutes what is commonly understood as the assertion of fatalism).

The Six Presuppositions

The argument relies, according to Taylor—and it has not been seriously disputed—on six presuppositions. These are:\(^1\)

\[(P1) \text{ Any proposition whatever is either true or, if not true, then false.}\]

(Taylor comments that '[t]his is simply the standard interpretation, tertium non datur, of the law of the excluded middle, usually symbolised } (p \lor \neg p), \text{ which is generally admitted to be a necessary truth'. As we have already seen in earlier chapters, it is really PB

---

\(^1\) Taylor (1968b: 223-4). Following the example of Segal (1981), and for the same reasons of clarity and ease of reference that he adduces, I have used Taylor's own words in a mixture of quotation and paraphrase. I have made some changes to the list but these are minimal, answering to cosmetic or stylistic rather than substantive considerations.
rather than LEM, but let that stand, since nothing important ultimately turns on it.)

(P2) If any state of affairs is sufficient for, though logically unrelated to, the occurrence of some further condition at the same or any other time, then the former cannot occur without the latter occurring also.

(P3) If the occurrence of any condition is necessary for, but logically unrelated to, the occurrence of some other condition at the same or any other time, then the latter cannot occur without the former occurring also.

(P4) If one condition or set of conditions is sufficient for (ensures) another, then the other is necessary (essential) for it, and conversely, if one condition or set of conditions is necessary (essential) for another, then that other is sufficient for (ensures) it. This is but a logical consequence of (P2) and (P3).

(P5) No agent can perform any given act if there is lacking, at the same or any other time, some condition necessary for the occurrence of that act.

(This follows simply from the idea of anything being essential for the accomplishment of something else.)

(P6) Time is not in and by itself 'efficacious'. This is just the idea that the mere passage of time does not augment or diminish the capacities of anything, and in particular, that it does not enhance or decrease an agent's powers or abilities.

It is Taylor's contention that (P1-6) generate a two stage proof of fatalism. The first stage establishes the truth of fatalism with regard to all past events; the second, establishes in a strictly analogous manner the truth of fatalism with regard to all future events. The sting consists in the fact that while we ordinarily acquiesce in the belief that we can do nothing at all to alter the past, most of us would recoil from the claim that we can do nothing to alter the future—although strictly speaking, as B-theorists are quick to point out, no one can alter the future either, all the while maintaining, if they are not fatalists, that we can affect it, in the sense of (helping to) bring it about.

One might question the need for the two stage approach. Robert Segal (1981), for example, in his explication of Taylor's argument concentrates solely on

---

2 Taylor says that this is '[a]nother and perhaps better way of saying the same thing' (1968b: 223). Likewise with respect to 'essential' for 'necessary condition'.
stage two, where he seeks to impugn \((A_0)\), his main target. But the first stage is essential to the attack on \((A)\)—and not just on polemical or rhetorical grounds (although of course it is extremely useful that he can point to seemingly identical situations, one of which accords perfectly with our everyday intuitions). Thus the first stage is meant to establish one half of what Taylor crucially wants us to see as a symmetrical argument. The argument of this first stage establishes fatalism about the past, with which few disagree; the second, fatalism about the future. If a fatalistic stance is appropriate in one case, why not in the other? 'A fatalist, in short,' says Taylor, 'thinks of the future in the manner in which we all think of the past.'

It is also not inappropriate at this point to remind ourselves that while, ontologically speaking, the B-theorist regards the future in the same manner as we all think about the past, she does however adopt an asymmetrical attitude towards our voluntaristic stance \textit{vis a vis} the future.

5.2 Fatalism about the Past

Taylor first presents the following scenario: I am about to read the morning newspaper. We are to assume that the prevailing conditions are such that if I read that a sea battle occurred yesterday (i.e., \(R\)), a sea battle occurred yesterday, and if I read that no sea battle occurred (i.e., \(\neg R\)), then no sea battle occurred yesterday. Assume further that there are two possibilities with respect to the battle itself:

1. \textit{No sea battle occurred.} In this case, a condition necessary for me to read that a sea battle occurred was absent (namely the occurrence of the sea battle yesterday). Why do we say this? We are assuming that \(\neg R\) is sufficient for the non-occurrence of the sea battle. So in this case a condition is absent (i.e., the occurrence of the sea battle) whose presence is necessary for the performance of the particular action (namely, \(R\)).

2. \textit{A sea battle occurred.} In this case, a condition necessary for my failing to read that a sea battle occurred is absent, namely, the non-occurrence of the sea battle. Once again, a situation exists where a condition is absent (this time the non-occurrence of a sea battle) whose presence is necessary for the performance of an act (namely, \(\neg R\)).

The crucial question is whether or not doing \(R\) is up to me; that is, is \((AR)\) true? For Taylor, it is quite obvious that \((AR)\) is not true:

\footnote{Taylor (1968b: 221).}
For if both these acts were equally within my power, that is, if it were up to me which one to do, then it would be up to me whether or not a naval battle has taken place, giving me a power over the past which I plainly do not possess...[W]hat sort of headline I see depends, among other things, on whether a naval battle took place yesterday, and that, in turn, is not up to me.\footnote{Taylor (1968b: 225).}

And this is entirely unsurprising since, as Taylor observes, we are all fatalists concerning the past.\footnote{I need hardly say of course that this isn't so. Presentists, for example, who do not believe in the reality of the past (e.g., Prior 1967) could not agree. Nor would someone like Lukasiewicz, who views the past as Aristotle viewed the future, and who believes that 'w[e should not treat the past differently from the future. If the only part of the future that is now real is that which is causally determined by the present instant ... then only those parts of the past are at present real which still continue to act by their effects today. Facts whose effects have disappeared altogether ... belong to the realm of possibility. One cannot say about them that they took place, but only that they were possible' (1967: 38).}

So much then for (AR); but what about (AO), the denial of which constitutes fatalism as it is commonly understood, namely, fatalism about the future?

5.3 Fatalism about the Future

We are to imagine the following scenario: an admiral is deliberating whether to issue an certain order O or to refrain from issuing O (i.e., issue \(-O\)). If the admiral issues O, a sea battle will occur tomorrow; if he refrains (i.e., issues \(-O\)), no sea battle will occur tomorrow. Once again there are two possible outcomes: either a sea battle will occur or it won't. Let SB be the proposition \textit{that a sea battle will take place tomorrow}, and \(\neg SB\) the proposition \textit{that no sea battle takes place tomorrow}. Taylor's case for the falsity of (AO) goes like this: \footnote{Taylor (1968b: 227).}

1. If SB is true, then it is not within the admiral's power to do \(-O\) (for in the case where SB is true, then there is, or will be, lacking a condition essential for his doing \(-O\), namely, the condition of there being no sea battle tomorrow).
2. But if \(\neg SB\) is true, then it is not within the admiral's power to do O (for a similar reason).
3. But either SB is true, or \(\neg SB\) is true.
Therefore, either it is not within the admiral's power to do O, or it is not within the admiral's power to do \( \neg O \).

But clearly the truth of (4) entails the falsity of \( A_O \). So if this argument is sound, then \( A_O \) is false.

Premise (3) is our old friend PB (or LEM, by Taylor's light). We shall agree for the moment not to challenge it.\(^7\) Premises (1) and (2) become then all-important. They are essentially the same premise (in that they do the same work) and thus stand or fall together. In order to see how they might be argued for, we need to make explicit the reasoning they employ. The consequents of (1) and (2), respectively, constitute the disjuncts of (4), so that each premise is in effect an argument for one of these disjuncts. The following sub-arguments are taken from Segal (1981):\(^8\)

**Argument for Premise (1)**

(5) Assume that SF is true.  
(6) If SF is true, then \( \neg \text{SF} \) is false.  
(7) \( \neg \text{SF} \) is false.  
(8) Doing \( \neg O \) is sufficient for \( \neg \text{SF} \).  
(9) \( \neg \text{SF} \) is necessary for \( \neg O \).  
(10) Therefore, the admiral cannot do \( \neg O \). 

**Argument for Premise (2)**

(11) Assume that \( \neg \text{SF} \) is true.  
(12) If \( \neg \text{SF} \) is true, then SF is false.  
(13) SF is false.  
(14) Doing O is sufficient for SF.  
(15) SF is necessary for O.  
(16) Therefore, the admiral cannot do O.

**The Conclusion**

The admiral lacks the requisite freedom in that it is not up to him what he is going to do with respect to O. Whichever action he does perform is the one he had to

---

\(^7\) Taylor originally thought that the conclusion might be evaded by giving up LEM for FCSPs (1968b: 228-9), until Cahn convinced him otherwise. See Cahn (1964: 305) and Taylor (1964b: 307).

\(^8\) Segal (1981: 370).
perform—he is not able to perform otherwise than he does. And since it is precisely this ability to do other than one in fact does that is generally taken to be the hallmark of a free action, it follows that the admiral is not free. This result affects not only our admiral: the case is readily generalisable, so that we can assert that whatever actions agents perform are the only ones that they possess the ability to perform. This goes beyond the mere tautology that whatever happens had to happen. The argument, if sound, has real bite. The problem for the fatalist, after all, has always been to find a version of fatalism weak enough to be plausible, yet sufficiently robust to be interesting.

The argument is intended to have considerable rhetorical force as well: in so far as we are reluctant to believe that reading now about an event that occurred yesterday can affect that event, and thereby find fatalism about the past plausible, we must also, Taylor would argue, make a similar finding with respect to fatalism about the future. Taylor is particularly adept at such parity arguments and here he trades on our willingness to grant the falsity of (AR) in order to press the falsity of the future tense version of (A), namely (AO).

5.4 Criticism

These arguments clearly stand or fall on the basis of (P4) and (P5). It is principally these features that distinguish condition-fatalism from the truth-fatalism we have examined thus far. (P4) validates the transformation of (8) into (9) and (14) into (15), via the principle to the effect that P is a sufficient condition for Q then Q is a necessary condition for P. Mark Bernstein has referred to this as a 'logically unimpeachable principle'.9 (P5), i.e., the claim that no one is able to perform an act in the absence of a condition necessary for its performance, then delivers the conclusion, in the first instance that the admiral is not able to refrain from issuing O, and in the second, that he is not able to issue O. Since these are the only two possible cases, the admiral will either be unable to issue O or unable to issue ¬O. Hence (AO), which asserts his ability to do both, will be false. Taylor expresses his confidence in (P5) thus: 'This is no law of logic, and in fact cannot be expressed even in the contemporary modal logics, but it is nonetheless manifestly true.'10 Most commentators have not been as sanguine about (P4) and (P5) as Bernstein and Taylor. In particular, critics have viewed the apparently highly plausible (P5) with much suspicion. The argument generated an extensive and often lively debate in the

9 Bernstein (1990: 270).
10 Quoted in van Inwagen (1983: 44).
The most subtle of Taylor's early critics was undoubtedly John Turk Saunders; Taylor's most prominent—indeed only defender, apart from himself—was Steven Cahn. I will be concerned with more recent criticism of Taylor and (P5) and refer to the earlier debate only where and when it is appropriate.

**Truth as a necessary condition**

Before we look at (P5), I wish to point to a common misconception concerning the role played by truth in Taylor's argument. Robert Segal claims explicitly that a necessary condition for performing a given act is the truth of the proposition to the effect that the agent does it, and that if it is false that that agent does it, then she cannot do it. If it's not true that a given event occurs tomorrow then one can't have issued an order today that would have been sufficient (in the obtaining circumstances) to bring about the event. For Taylor then, according to Segal, one of the necessary conditions for the occurrence of a seafight tomorrow is the truth of the proposition that a seafight will occur tomorrow. If this is right, then it looks like a problem, for is truth the sort of thing that can be a necessary condition for the occurrence of an action, event or state of affairs? We can readily see why it is in a critic's interests to read Taylor this way: Segal implies that this is not the sense in which one would or should be inclined to agree with Taylor that (P5) is eminently plausible, since Taylor does not rely on the sense of it that Segal regards as generally plausible (i.e., one based on considerations of general know-how or ability). Storrs McCall makes a similar charge:

Taylor argues that no one can do anything, a necessary condition for which is lacking. If the necessary condition that is lacking is a presently obtaining empirical state of affairs, then Taylor's principle is correct. No

---

11 Many of his critics accused Taylor of committing a modal error; Moyal believes that there is a 'general consensus' that Taylor is guilty of a 'misplaced modal operator' (1985: 104). But in this he was only following in the footsteps of Abelson (1963: 95) who had also claimed that Taylor illicitly transferred the logical necessity of the disjunction from entire disjunction to individual disjuncts. Cahn defends him, pointing out that he argues if A then necessarily B and if ¬A then necessarily ¬B (and further that, in putting his case, Abelson has committed a similar error, in reverse, to that of which he accuses Taylor) (1967: 100).


one can saw a plank in half without a saw. But if what is lacking is the truth of a future-tense proposition, then Taylor's principle is not correct. Truth and falsehood are not empirical states of affairs. What is possible or not possible for us to do is independent of what we will do, and hence of the truth or falsehood of the corresponding propositions. We can, therefore, do things we will not in fact do. (McCall 1994: 16 fn.20)

I do not believe that Taylor makes this mistake of regarding truth as a necessary condition. The necessary condition is the event the true proposition describes, not truth itself. He seems to say quite explicitly that the condition in question is the event of the naval battle, or the event of the non-battle. Look again at Premise (1) in the argument for the falsity of \( (A_0) \). He states: '...there is, or will be, lacking a condition essential for his doing \( -O \), namely, the condition of there being no sea battle tomorrow'. I see no mention here of truth. If it were the case that Taylor relied on truth in his argument, it would tend to collapse back into the type of truth fatalism that Aristotle considered. It retains its integrity as a distinct argument only to the extent that it relies on a central role being played by the events themselves.

This error explains why many commentators conflate his two arguments for fatalism, either explicitly or implicitly. It is readily done if you take truth to be one of the necessary conditions for an event's occurrence. McCall, for example, seems to think that in the context where Taylor uses (P5), he is still arguing for truth fatalism, as he had done in his story of Osmo. Even Gaskin, in his (1994) article, seems to run the two together in the course of his exposition and commentary. It is true that we do not ordinarily regard true propositions as relevant factors in determining what a person can and cannot do. But truth is not the central issue here: it serves only to point in the direction of the existing events that comprise the relevant conditions. And these events are empirical states of affairs, and as such, determinate features of reality which can play the role Taylor has set aside for them.

5.5 The Truth of Presupposition 5

Taylor's argument depends on many other assumptions (including Bernstein's logically unimpeachable principle), but the crucial one—and the one that has been most often questioned—is (P5). Is (P5) true? The truth or not of (P5) is the main issue, and not just because of the central place it occupies in Taylor's explicit argument. Van Inwagen neatly sums it up:

The crucial premiss in Taylor's argument is Principle (A) [i.e., (P5)]. If we accept this principle, then we can derive fatalism by a much simpler argument than Taylor's. For it is obvious that whenever I am not performing a certain act, then there is absent a condition necessary for my performing it: namely, my performing it. And, of course, whenever I am performing a certain act, there is absent a certain condition necessary for my not performing it: my not performing it. Thus, Principle (A) leads directly to the collapse of the distinction between what one does and what one can do: one who accepts Principle (A) has already got fatalism in his pocket and need not shop for it in Taylor's elaborate naval bazaar, diverting though the wares offered there may be. (1983: 45)

Van Inwagen's suggestion that (P5) alone would suffice to establish the fatalist conclusion is more to the point than his somewhat off-hand manner might tend to suggest. If it is true, as I having been arguing, that existing features of reality are the stuff of which fatalist necessity is made, then it will be the case that fatalism can be achieved more directly than via the somewhat roundabout route through the language of necessary and sufficient conditions. It is easy to see what motivates the fatalist to employ such methods. If it can be suggested that somehow I am constrained by some other feature of the world that is inaccessible to me, then the force of the claim that future events are unavoidable will be more keenly felt. It was precisely the same type of belief that motivated the decision to talk in terms of prior truth about future events.

It is one thing to be told that, according to a particular view of temporal reality, two alternatives are not really open to me with respect to some future decision I have to make; it is quite another to be told that the reason for this narrowing down of possible outcomes lies elsewhere, either safely in the untouchable past or else, as in this case, tucked up in the future, yet apparently none the less potent in virtue of a somewhat mysterious relation of conditionship obtaining between the choice I must attempt to make and that still more future necessary condition. This feeling is reinforced by whatever residual A-theory intuitions I may possess, intuitions, which lurking deep in one's metaphysical heartland, lead one periodically to think of certain events as being less fixed than others. Thus I imagine that my choice now is somehow more fluid than last year's choices. But if all events are equally settled in reality, then this feeling is a misleading one. A sense of openness about the future may be a psychological necessity, but that doesn't mean that it is objectively grounded.
5.6 The Ambiguity of Presupposition 5

In any case, van Inwagen objects to what he sees as a crucial ambiguity residing in (P5). As a result, it isn't possible to say whether it is true or not without first disambiguating it. One interpretation of (P5), he claims, renders it obviously true, but in a way that gives no support to the fatalist, while the other interpretation, although it supports the fatalist, seems clearly false. Slightly modified, these are van Inwagen's two readings:

\[(\alpha)\] No agent is able to: \textit{perform a given act in the absence of a condition necessary for its accomplishment.} \\

\[(\beta)\] No agent is able, in the absence of a condition necessary for its accomplishment, to: \textit{perform a given act.} \\

In other words, it is the difference between 'I cannot do X in the absence of a necessary condition for X' and 'I cannot do X in the absence of a necessary condition for X.' Let's take a concrete example: imagine a I am a pole-vaulter locked in a room with a 10' ceiling. It is, we will assume, impossible to pole-vault in such a room—you require at least 20' of clearance. So it is the difference between 'I cannot pole-vault in a room with a 10' ceiling' and 'I cannot in a room with a 10' ceiling: pole-vault.' Before we can assess the implications of these two readings, we need to understand what sense of 'necessary condition' is being used in them.

\textit{The nature of the necessary condition in (P5)}

What then is the nature of this necessary condition? That is to say, when Taylor speaks in P2-4 of necessary and sufficient conditions, does he mean \textit{logically} necessary and sufficient, \textit{causally} necessary and sufficient, or something entirely other? This was one of the earliest points seized upon by Taylor's critics and it has remained in the fore of the discussion. One of his earliest critics, Abelson (1963), claimed that Taylor confuses causally necessary conditions with logical necessity. This is despite Taylor's denying that he introduces causation into the argument at all, and his denying explicitly in P2 and P3 that his brand of sufficiency is logical. A recent critic, Keith Seddon, argues that Taylor equivocates between this logical sufficiency, which was supposedly ruled out, and 'Taylor's undefined sufficiency',

\[15\] van Inwagen (1983: 45).

\[16\] Abelson 1963: 94-5).

\[17\] Taylor (1968b: 228): 'our problem has been formulated without any reference to causation'.

which Seddon claims is just causal sufficiency. After all, how else should you interpret the claim that 'ingestion of cyanide ensures, or is sufficient for death', as Taylor says in explicating P2. Bernstein says that 'Taylor explicitly wishes us to' read 'necessary' as meaning 'causally necessary', and van Inwagen reports, 'Taylor has told me that he intended Principle (A) [i.e., (PS)] to apply only in the case of conditions causally but not logically necessary for one's acts.'

Let's agree to take Taylor to mean causal, rather than logical, sufficiency and necessity. But why is it important? Seddon argues that the following principle (namely, Bernstein's 'logically unimpeachable principle')

(LUP) If x is sufficient for y, then y is necessary for x

is fine for notions of logical sufficiency and necessity, but that it is wrongly applied by Taylor to his 'special undefined sort of sufficiency', which Seddon maintains to be a causal relationship, albeit by stealth. Seddon claims that if so applied it would mean that the ingestion of cyanide being causally sufficient for death would entail that death is causally necessary for the ingestion of cyanide. And this just seems wrong: 'The idea that a future circumstance can be causally necessary for something that happens now is difficult to understand.' Now Taylor has repeatedly pointed out that one of the very things he is saying is just that the role played by necessary conditions that are lacking in the past is an obvious one but that '[i]t is less obvious when one considers necessary conditions which are lacking in the future'. So perhaps the strangeness of this can be overplayed. Taylor is, after all, attempting a

21 van Inwagen (1983: 229 n. 20). Van Inwagen thinks that this restriction is pointless and that Taylor ought to extend (P5) to cover logical conditions - 'Surely if it's plausible to suppose that I can't do a thing in the absence of a condition causally necessary for my doing it, then it's even more plausible to suppose that I can't do a thing in the absence of a condition logically necessary for my doing it' (1983: 229 n. 20).
24 Taylor (1962: 25-6).
certain degree of conceptual revision. In any case, I think ultimately that this isn't as significant as might be thought at first. It seems to me that what he is really grappling with is the idea that all events, in so far as they commonly partake of full reality, whether in the past of the future, are related in ways that perhaps the concepts of logical and causal sufficiency can't precisely capture. The upshot will be that the debate between Taylor and his critics will eventually come down to a disagreement over what constitutes human ability.

The (α) reading of (P5)

But first we have to return to our example and see what it looks like (i) on an interpretation that takes the condition to be logically necessary, and (ii) when it is taken to be causally sufficient. For the time being, I will only be considering version (α) of P5.

(i) Logically necessary condition. I cannot perform an action in the absence of that action. For example, I cannot pole-vault at \( t_n \) in the absence of my pole-vaulting at \( t_n \). That is just to say that I can't pole-vault and not pole-vault at the same time. But, as Bernstein quite rightly observes, this makes (α) trivially true. This has a profound effect on the debate over arguments like:

17. No one can: X in the absence of X-ing.
18. \( \neg X \)
19. Therefore, no one can X.

The first premise (17) is, as Bernstein points out, 'merely an instance of the claim that one cannot perform the (logically) unperformable' and hence a tautology. It thus drops out, and we are left with the apparently invalid inference

18. \( \neg X \)
19. Therefore, no one can X

until we recall that it is just the fatalist's contention that this type of inference is valid, and it cannot just be dismissed as invalid, on pain of begging the question against him. Of course, neither can Taylor simply assume it to be valid without also being guilty of begging the question. What I would want to say here is that the

---

26 Bernstein (1990: 272).
deadlock can be broken by attending to the theory of time that affirms the full reality of the future.

The anti-fatalist who wants to claim that I can pole-vault, even though I'm not actually doing so now, might interject at this point. She does not claim the ability to pole-vault and not pole-vault at the same time—just the unexercised, yet retained ability to pole-vault when not actually pole-vaulting. But you have to ask, what would it be actually to exercise that ability? It would be to perform the logically unperformable—pole-vaulting and not pole-vaulting. The appropriate context in which to make this claim is the causal one, to which we now turn.

(ii) Causally necessary condition. On this view of (α), Taylor is understood to mean that the relevant condition that figures in instances of (P5) will be a causally necessary one. The trouble with this, however, is that taking (α) as trading on causally necessary conditions seems to go too far the other way: while it is no longer trivial, it is now questionable whether it is actually true. Bernstein feels that we ultimately will come down again to a question-begging disagreement with the anti-fatalist over the truth of (α).

The fatalist and the anti-fatalist disagree over notions of ability. Both will agree that I cannot pole-vault in a room with a 10' ceiling. But they will disagree over what this means. I lack the power to do it in the circumstances, and no amount of general ability or technical know-how is going to make up for that. The anti-fatalist wants to say that from my lacking a condition (causally) necessary for my pole-vaulting, all that follows it that I am do not pole-vault, but not that I am unable to. 27 Not only that, I can pole-vault even when not pole-vaulting, if this is now interpreted causally, viz., as shorthand for 'I can pole-vault, even when not actually pole-vaulting because the prevailing circumstances aren't currently propitious (e.g., the ceiling's too low).'

But what does the fatalist mean by the statement that I am unable to pole-vault in the circumstances? Taylor makes it quite clear that one's not being able to

27 This point was made by J.T. Saunders, the sharpest of Taylor’s early critics, and we will encounter it again in numerous other hands (1962a: 1); cf, Taylor (1962: 25), where he formulates Saunders' objection in these terms. Saunders further claims that '[t]he occurrence of a naval battle on the morrow is a necessary condition of O but not of the ability to issue O' (1962a:2). But if this means that by virtue of being a necessary condition of O, it has effectively fixed it, then it doesn't matter what he says about the ability to issue (or not issue) O any more; it will be issued.
do something does not mean that it is logically impossible to do that thing: 'the fatalist argument has nothing to do with impossibility in those senses familiar to logic. It has to do with unavoidability. It is, in other words, concerned with human abilities'.

Therefore it just means that issuing the order is not within his power. But it is not a matter of denying any general ability to do something like pole-vaulting. Rather, for Taylor it means just that an act is not within one's power unless all the necessary conditions for its accomplishment are satisfied. The key notion of 'within one's power' is not to be conflated with 'can' or 'is able' or even 'possible'. Taylor sees this confusion on the part of his critics as their 'fundamental error'. For their part, they see him as equivocating in the use of the term 'can'.

Taylor's mistake, in Saunders's view, is to take 'can' to mean 'has the power to'.

Taylor is very wary of the ordinary sense of ability. This ordinary sense involves concepts such as skill, know-how, having the requisite physical attributes, and so on. This reason for his wariness is his belief that if you press this sense too hard, 'then we need no longer be fatalists about the past, for we will then have the ability to do things that are sufficient for the occurrences in the past of things which did not in fact occur—for instance, to make a battle occur yesterday which did not occur'.

Take the example of someone who actually claims to have the power to change the past. That such claims have been made, despite the widespread feeling that such action is impossible, shows that Taylor's worry was not an idle concern.

Peter Damian is supposed to have claimed for God the ability to make undone the past in the sense that God's omnipotence entailed that he could make it that Adam never existed (and not just in the sense that today Adam existed and tomorrow he did not). There is some dispute over exactly what Damian's claim was, and it is relevant to this discussion. McArthur and Slattery (1974) see it as being central to Damian's position 'that strictly speaking there are no past facts'. Remnant (1978), however, claims that 'Peter never suggests that God actually does tamper with the past—merely that he could if he wanted to'.

Unsurprisingly, on closer examination, this latter interpretation of Damian mirrors the sort of ability that Taylor warned about.

---

29 Taylor (1963: 497).
31 Taylor (1962: 26).
33 Remnant (1978: 265).
We can contrast two senses of power: in the first sense, the power to perform a particular act under given circumstances and, in the second, a general sense of power to perform acts of a certain kind. The distinction can be exemplified thus: if I am a skilled pole-vaulter but am presently locked in a room with a 10' ceiling, then I lack the power to pole-vault. But if it is the case, as we are assuming, that I possess the general power or ability to pole-vault, then there is, on the second sense of power, nothing in the circumstances which prevents (in a causal sense) or precludes (in a logical sense) me from pole-vaulting. As Hasker says,

There is a way of speaking about powers according to which a person's powers are thought of not primarily as powers to perform a specific action on a specific occasion, but rather as general abilities that one has and that remain more or less constant although the possibilities of their exercise come and go.34

Hasker further maintains that 'I may perfectly well have a power of this sort to do something even though it is either logically or causally impossible that I exercise the power under the circumstances that obtain at a particular time'.35 The context for this is a discussion, by Hasker, of a modern day Damian who claimed to be able to alter or change the past, and not just in the sense of being able to influence it or help to bring it about, but rather in the sense of being able to take something in the past that was the case and make it, not merely no longer the case, but never even having been the case. That is, he is talking about changing the past in the sense of now making something in the past otherwise than it was.

This seems to have been precisely Taylor's worry—that if you allow this general sense of ability to dictate your assessments of what is within someone's power to perform, then it can lead to such apparently absurd claims. Any analysis that has these consequences must be faulty. At one level, Taylor and his critics disagree about how one ought properly to define ability, but at a deeper level, Taylor is making the valid point that, if you want to avoid such bizarre claims, you should attend to what agents actually can and cannot do in the existing circumstances. After all, these circumstances are always there, providing the context within which one must exercise whatever power one has. This notion of context-specific power captures the essence of the second version of (P5).

The (β) reading of (P5)

It is this version that van Inwagen claims would deliver fatalism, were it only true, but there is no reason to think that it is so.\(^{36}\) The (β) reading of (P5) has the form: 'I cannot do X in the absence of a necessary condition for X.' It generates the instance that says that 'I cannot in a room with a 10' ceiling: pole-vault'. The 'ability operator' has been expanded and thereby relativised to the relevant circumstances in any particular case. If the example under consideration is one that is interpreted in terms of a logically necessary condition, this operator would be 'No agent can X in the absence of X...'; if it is interpreted causally, it would be 'No agent in the absence of a causally necessary condition for X can...'. If this relativisation is permitted, then I can't pole-vault when not actually pole-vaulting (a generalised sense of ability notwithstanding), nor can I pole-vault when not in the right sort of arena for such acts.

But in order to impart legitimacy to the application of (β), Taylor needs to specify the grounds on which it is proper to pack the prevailing circumstances into the first part of 'No one is able to: do X'. When would it be appropriate to so relativise ability claims in this way, so that 'No one is able to....' becomes 'No one is able in the absence of a necessary condition for X...'? I would argue that it is entirely appropriate to do so in the context of a B-theory ontology. This is because the event of not pole-vaulting at \(t_n\) is unalterably part of the world (just as much, and in precisely the same manner, as is the case with any past event). Therefore, if I do not pole-vault at \(t_n\), then I can't pole-vault at \(t_n\), general ability to pole-vault notwithstanding. The crucial point is thus not so much to decide between whether Taylor is making use of logical or causal conditions, but rather whether this manoeuvre is appropriate. For if it is, then van Inwagen's short cut to fatalism will be a reality.

Thus far I have been dealing with various arguments (a) to the effect that the B-theory is fatalistic, and (b) that have been advanced on behalf of fatalism. In Part Three, I want to examine the notion that the assumption of fatalism carries explanatory power in what are sometimes felt to be otherwise intractable domains.

\(^{36}\) van Inwagen (1983: 49).
PART THREE

OLD WINE IN NEW BOTTLES?
CHAPTER 6
THE PARADOXES OF TIME TRAVEL

Emitting another sigh, Stoker held his head and said, 'I don’t understand all this [i.e., time travel]. Yet I’ve umpired cricket matches without difficulty.'

Brian Aldiss
Dracula Unbound

All is now secure and fast;
Not the gods can shake the Past;
Flies-to the adamantine door
Bolted down forevermore.
None can re-enter there, —
No thief so politic,
No Satan with a royal trick
Steal in by window, chink, or hole,
To bind or unbind, add what lacked,
Insert a leaf, forge a name,
New-face or finish what is packed,
Alter or mend eternal Fact.

Ralph Waldo Emerson,
The Past

6.1 Introduction

In this chapter, I examine three related but nonetheless distinguishable instances where the possibility of time travel appears to engender a sense of paradox. It is my contention that the insights embodied in the fatalist stance I have developed can be successfully applied to the resolution of these difficulties. My purpose therefore, in this and the next two chapters, is to use the time travel scenario (and later scenarios involving backward causation and Newcomb problems) as illustrative of the explanatory role which fatalism either can or cannot play. I believe that it is definitely indicated in the case of time travel, a possible option in the retro-causal case, and unlikely to feature prominently in the case of the difficulties peculiar to the Newcomb problem case. Nevertheless, each case will provide significant insights. (Of course, if fatalism is true, then we could expect it to apply universally to these and other situations, but for now I am concentrating on the degree of plausibility each situation confers on the fatalist approach to explaining them.)

The present example—time travel—is crucial: in fact, in one sense it forms the core of this work in that the problem of circular time which provides the physical context for the discussion has a great deal to tell us about the view of reality that is equally committed to the existence of past, present and future events. This is
because the appropriate metaphor for circular time in a time travel universe is not, as might first be suspected, a closed curve where the ends have fused to form a smooth circle, but rather an unjoined loop in a line that curves back on and over itself. This latter image represents merely a topological deformation of the straight line image typically used to represent the B-theory view of temporal reality. In such contexts, as we shall see, the ultimate conceptual responsibility is for consistency, both causal and logical. Ontological fatalism is able to provide a metaphysical explanation for the existence and particular nature of these constraints.

I will first examine the so-called 'Grandfather Paradox' of time travel, focussing on David Lewis's development of the now almost standard response, and then consider essentially analogous paradoxes not involving human agency. From there, I go on to discuss an additional and seemingly very powerful objection to the possibility of time travel which Graham Nerlich has raised and which purports to resist a Lewis-type resolution to the Grandfather Paradox.

Time travel is usually discussed within the context of a B-theory, which if nothing else provides a destination for the time traveller. Time travel thus regarded looks quite impossible given an A-theory of time. In addition, it would seem to require a 4-dimensional view of persons and objects—if for no other reason than that time travel would seem to entail that certain objects (e.g., the time traveller) be bilocated, i.e., exist at two different places at the same time. David Lewis, for example, states that a time traveller talking to himself on the phone 'looks for all the world like two different people talking to each other'.1 By the hypothesis of time travel, I shall understand the notion of some agent travelling from the future back to the past. The world line of the agent in question bends backward relative to those of the other non-time-travelling individuals with whom he had hitherto been associated, i.e., with the world lines of the earth and its inhabitants. It involves, as Lewis has observed, 'a discrepancy between time and time'.2 I will later spell out exactly what this means. But, first, is there any empirical backing for time travel?

The Physical Possibility of Time Travel

In 1949 Kurt Gödel found solutions to the field equations of General Relativity which permit closed timelike lines or curves (hereafter CTCs) to exist in spacetime. Since a timelike line is any spacetime path along which it is possible that there be a

---

1 Lewis (1986: 72).
2 Lewis (1986: 67).
causal chain, these solutions permit the existence of *closed causal chains*. Gödel's cosmological models have the property that there is a closed, future-directed timelike curve through every point of spacetime and thus 'by making a round trip on a rocket ship in a sufficiently wide curve, it is possible in these worlds to travel into any region of the past, present, and future and back again, exactly as it is possible in other worlds to travel to distant parts of space'. It is important to realise that the traveller's worldline would always be oriented toward the locally defined future, yet the curved global structure of spacetime—the way the local regions are pieced together—would permit his arrival home to precede his departure. It is no less important to note that in Gödel's universe it is not possible to define a universal or cosmic time because all the local times of observers which are associated with the mean motion of matter cannot be fitted together into one world time.

It seems then that General Relativistic situations involving the possibility of causality violations can be formulated. Dwyer has noted that 'such time travel will of course involve backward causation, for, from any event \(x\), a signal emanating from \(x\) can influence certain events which lie in \(x\)'s chronological past'. That is to say, it seems possible that an event that lies in an agent's chronological past could also lie in his causal future and thus be subject to his influence. Many physicists find such consequences to be objectionable since endowing agents with such abilities appears to violate some of our most fundamental notions concerning how the world operates. As a result, they think it entirely proper to impose on physically acceptable spacetimes the requirement that they not be characterised by such causality violations. Yet it is not at all obvious that these violations, although admittedly counter-intuitive, ought to be rejected *a priori*. And while it seems empirically to be the case that this model does not in fact describe the real universe, this does not diminish its nomological possibility. Earman (1967), Stein (1970) and Weingard (1979) have all defended the theoretical possibility of Gödel's model. It seems to be at least broadly possible that time travel along Gödelian lines might exist.

Other possible models exist: for example, Kip Thorne and others have shown how the two ends of a traversable "wormhole" through spacetime could be positioned in order to form a CTC; J. Richard Gott has calculated that two infinitely long cosmic strings (albeit merely theoretical entities) moving very rapidly past each other would generate CTCs. There is also Tipler's 'rotating cylinders' idea, which is similar to Gödel's approach except that it uses a much smaller mass than the entire

---


4 Dwyer (1978: 21).
Stephen Hawking firmly believes that nature abhors time machines like these. He has ventured a *chronology protection conjecture*, which says that the laws of physics do not allow time machines:

Whenever one tries to make a time machine, and no matter what kind of device one uses in one's attempt (a wormhole, a spinning cylinder, a "cosmic string", or whatever), just before one's device becomes a time machine, a beam of vacuum fluctuations will circulate through the device and destroy it.6

Nature is thereby able to 'keep the world safe for historians'.7

But we do not have to look exhaustively through these possibilities in order to find one that is empirically plausible. It is enough for our purposes that current theory does not rule out the possibility of time travel. My concern lies elsewhere: problems of *logical consistency* arise when we start to consider what a time-travelling agent could or could not do in the past.

A PARADOX INVOLVING HUMAN AGENCY

6.2 The Grandfather Paradox

The notion that the possibility of time travel leads to paradoxical situations involving either one's earlier self or one's ancestors turns up repeatedly not just in the philosophical literature but also in works of science fiction and popular science. The basic scenario is quite familiar:

In the classic inconsistent time-travel paradox, a man travels back in time to kill his own grandfather, in which case he ceases to exist. But if he never existed, how can he kill his own grandfather? The logical impossibility of such a situation is the strongest argument against time travel.8

Whatever has already happened cannot now be undone. But if someone could return to some time in the past, he would be able to bring about a

---

5 For all of these proposals, see Thorne (1994: Chapter 14).
state of affairs that, as a matter of historical fact, did not occur at that time. This is a contradiction. Therefore we can infer either that such time travel is impossible or that time travelers are constrained by mysterious forces that conspire to prevent them from bringing about such contradictions.9

As is clear from these two passages, there are two senses in which the paradox arises: (i) because an event would both occur and not occur (logical inconsistency), and (ii) because two events would occur which are causally inconsistent. These two types of inconsistency go right to the core of the time travel paradox.

In his now classic paper, 'The Paradoxes of Time Travel', David Lewis considers an argument purporting to demonstrate the impossibility of time travel:

(1) A time traveller visiting the past would be as able as any agent to do things that, if he did them, would change the past.
(2) A time traveller visiting the past could not change the past because changing the past is impossible.
(3) Therefore, a time traveller visiting the past both could and could not do things that would change the past.
(4) This involves a contradiction.
(5) Therefore, there cannot possibly be such a time traveller.10

In another seminal philosophical discussion of time travel, 'Who was Dr Who's Father?', Murray Macbeath offers an almost identical version of this argument:

More briefly the objector could say that, if time travel were possible, it would be possible for you to kill your grandfather when he was five years old (the obvious assumption being taken for granted). Clearly it is not possible that you should kill your grandfather when he was five—otherwise you would not exist to do the killing. And so we have a contradiction: on the assumption that time travel is possible, you both could and could not kill your five-year-old grandfather. And the assumption that time travel is possible must, therefore, be rejected.11

This similarity is not surprising given Macbeath's unreserved endorsement of Lewis's response to this argument, but what is of interest is the fact that they seem to regard

---

10 I have reconstructed this argument from Lewis (1986: 75).
this as an important and widely held argument purporting to demonstrate the logical impossibility of time travel. The arguments, however, are not exactly the same: consider the reasons given for the inability claims in each case. There are, broadly, two distinct types of problem associated with the time traveller's attempt to kill his grandfather. Lewis relies (primarily) upon the idea that success would involve changing the past, given the assumption that it is a fact that the intended victim was not killed at the time in question, and on the impossibility of changing the past. Macbeath ascribes the inability to the fact that, had the time traveller been successful, he would thereby have precluded his own very existence. Lewis actually mentions both reasons: the factuality of the past difficulty as well as the causal one. It is precisely here, as we shall see, that the logical and causal inconsistencies threaten to enter. Given all this, ought we to conclude that the time traveller cannot kill his grandfather? And if he can't, what constraints him? The principal task for the defender of the possibility of time travel is to explain how it is possible for the first disjunct of the conclusion Horwich describes above to be false while avoiding the mysterious, quasi-fatalistic compulsion of the second. I want to examine this question primarily through a detailed examination of Lewis's version of the argument against time travel. Later I will go on to examine alleged paradoxes that do not involve the problematic notion of free agency which Lewis and others are concerned to retain along with the logical possibility of time travel.

Lewis asks us to consider the case of Tim, who hates his grandfather and wishes to kill him. Using money inherited from Grandfather [hereafter 'GF'], Tim constructs a time machine and returns to 1920. His intention is to kill GF one day in 1921. We are to take it as an assumed fact about the past that GF in fact died in 1957.

The gist of the Lewis/Macbeath argument is that it seems to be the case that a time traveller both can and cannot do certain things, and therefore since the assumption of time travel leads to contradiction, time travel is impossible. Lewis's strategy in response to his version of the argument is to deny the fact of contradiction, claiming that it is merely apparent due to an equivocal use of 'can':

We have this seeming contradiction: "Tim doesn't, but can, because he has what it takes" versus "Tim doesn't, and can't, because it's logically impossible to change the past." I reply that there is no contradiction. Both conclusions are true, and for the reasons given. They are compatible because "can" is equivocal.\[12\]

---

\[12\] Lewis (1986: 77).
That is to say, in one sense our time traveller can and in another he can't do that thing which, were he to do it, would change the past. That thing is to kill his grandfather. You must ultimately choose between 'can' and 'can't'. But to choose 'can't' seems an admission of fatalism.

At least two other responses are possible: either there is no good argument for the cannot case and compelling reasons to assent to the can case, or the argument for the can case is unsound and that for the cannot case is compelling. No-one seriously argues for the first option; the second however is another matter, as we shall see. I will argue that this second option ought to be adopted and that, generally, we should not be too precious about the retention of capabilities involving free will. In any case, Lewis's treatment needs first to be unpacked more thoroughly via a more extensive examination of the arguments to the effect that Tim can and that Tim cannot kill GF.

*Lewis's Version of the 'Can' Case*

Lewis's argument appeals (in (1)) to widely held and strongly felt intuitions about the abilities of agents and seeks to establish that Tim shares these abilities. Tim 'has what it takes' in terms of motive, means and opportunity, to adopt the murder mystery vernacular. He has everything going for him: gun, training, good weather, a secure hiding place and so on—it's easy to imagine the story. 'In short, Tim is as much able to kill Grandfather as anyone ever is to kill anyone'.13 Mention here of 'anyone' directs us to Lewis's device of the 'Tom analogy'. This analogy is meant to render more plausible the can case, and at first glance it seems a very powerful means of doing this. Tom is an ordinary, non-time-travelling agent out to kill GF's partner. Tom's situation with respect to motive, means and opportunity is exactly analogous to Tim's and we have powerful intuitions that tempt us to admit that Tom can kill his intended victim. But '... Tim has everything going for him that Tom does'.14 For all these reasons Lewis concludes that Tim can kill GF, and since Tim is a time traveller,

\[ (6) \text{ Time travel entails that Tim can kill GF} \]

13 Lewis (1986: 75).
14 Lewis (1986: 75).
in the sense that if time travel were possible then it looks like it would be possible for Tim to kill GF. The problem is that this would involve changing the past; but this is universally held to be impossible. The worry then is that time travel involves changing the past:

(6) Time travel entails that Tim can kill GF.
(7) But killing GF involves changing the past.
(8) Therefore, time travel involves changing the past.

However, as it stands there is clearly something wrong with this argument. The consequent of (6) is not picked up in the antecedent of (7), which refers, more strongly, to the actual killing of GF. A natural modification of (7) to remedy this would be:

(7') The ability of Tim to kill GF involves changing the past.

But (7') is not plausible. If I father a child then this obviously involves my becoming a parent (at least in the biological sense), but my acknowledged ability to father a child does not entail that I actually do so. If the can case is right, then it is quite proper to imagine scenarios where Tim's ability to kill GF is conjoined with his failure to do so. So there is no reason to think (7') true, yet it is required to get to the conclusion (8), (7) being insufficient as it stands. We therefore feel entitled to question (8).

Now, as it happens, (8) itself actually functions in what is perhaps the most widespread argument against time travel. In L. Dwyer's version, it runs as follows:

(8) Time travel involves changing the past.
(9) But it is logically impossible to change the past.
(10) Therefore, time travel is logically impossible.

The defender of time travel typically resists this argument by denying the first premise, namely, (8). Dwyer and others point out that there could well be time travel without any changing of the past. This is entirely plausible in light of the failure noted above of the argument involving (6), (7) and (8). Nevertheless let's look more closely at (8) in order to see why rejecting it makes sense.

15 Dwyer (1975: 341).
What does (8) actually assert? It appears to admit of two possible interpretations:

(8a) Time travel involves (actually) changing the past.
(8b) Time travel involves the possibility of changing the past.\(^{16}\)

If the sense expressed in (8a) is intended, then (8) is pretty obviously false. We can readily imagine circumstances where time travel would occur and yet changing the past might not. And we could imagine this happening in every instance. Indeed this is just what Lewis and others would claim occurs in cases like Tim’s: the consistency of the story is maintained because Tim somehow fails in his attempt to change the past. Repeated failure of this sort may well be implausible, but this does not detract from the possibility that this is sometimes the outcome. Therefore, it cannot be true that if time travel exists then the past will be changed.

We're left then with (8b). And this seems to be exactly what we want. Time travel has resulted in Tim being in a position where he is able to do something which he will not in fact do, but which, were he to do it, this would result in changing the past. So, time travel entails the possibility of changing the past, but this is fine since it's a possibility that will never in fact be exercised. Let's leave to one side for the moment the obvious objection that there's not a lot to be said for an ability which one only possesses on the proviso that it never actually be exercised. We'll accept the point that '[Tim's] failure by no means proves that he was not really able to kill Grandfather'.\(^{17}\) Our problem is that time travel seems to have the absurd consequence that Tim can bring it about not only that he never existed himself or had the means to travel to the past (thus incurring causal inconsistency), but more

\(^{16}\) We could actually have made the same point had we changed (7') such that it read

(7") But the ability to kill GF involves the possibility of changing the past.

This would have allowed us to conclude (8b).

\(^{17}\) Lewis (1986: 76). In a roughly analogous case involving tachyonic signalling from the future, Paul Fitzgerald comes to a similar conclusion, his agent reasoning thus: "'It is in my power to kill Rameses if I choose. For my killing or not killing Rameses depends crucially on what I am inclined to do. Had I wanted to kill Rameses, then I would succeed in killing him. Of course, in that case I would not have this reliable message from the future telling me that he is alive and well. So, I know that I won't kill Rameses." ... The paradox is resolved. Nothing is preventing you from killing Rameses; you have the ability, the opportunity, and the wherewithal. If you want to and try, you'll succeed. The fact that you know by tachyon message that you won't kill him does not by itself testify to any limitation of your freedom to kill him' (1972: 433).
importantly that an event which, by assumption, did in fact occur, 'does' not occur (thereby incurring logical inconsistency). But anything that has absurd consequences is itself incoherent. The alleged conclusion to be drawn is that, as a result, time travel is not possible. The defender of time travel responds by pointing out that we can escape this conclusion so long as this ability is not exercised. But this really misses the point. The problem is not that one has an ability that would have absurd consequences were one to exercise it, but that's not a problem because it turns out that it isn't in fact exercised. The absurdity lies essentially in the postulation of the ability itself. And generally, if the ability postulated is one that, were it to be really exercised, a contradiction would ensue, then surely we are entitled to deny to the agent in question the claimed ability. Tim is to perform an act that is inherently self-contradictory. Time travel merely puts Tim in the position where he is able to do something which, were he to do it, would have absurd consequences. But it is the postulated ability itself—the ability to change the past—that is incoherent. This would be absurd even in contexts unrelated to time travel. There is obviously something wrong somewhere, but it is premature to assume that the problem lies with time travel. And why assume that you have dispensed with all the problems just because you have shown that the possibility of time travel is compatible with an unexercised ability? This could all be true and yet problems still remain in the guise of the alleged ability itself.

In order to see this, consider again the case of a person in a room with only one (apparent) exit. In truth, the 'door' is not a real door at all and therefore wholly impassable. Imagine our prisoner never attempts to exit via this ersatz door. Applying Lewis's principle, we might want to conclude that her failure to exit the room via this 'door' doesn't prove that she cannot, or is not able to do so. Quite true. But in fact she cannot leave through this door. She is not really able to leave. This fact is quite compatible with the truth of the claim that it doesn't follow from the fact of failure that she is unable. It doesn't follow from this fact; rather it follows from certain ontological facts about the room and its 'door', i.e., about the state of the world. Given this situation, it is clearly odd to claim that she has the ability to exit through an impassable doorway. There is nothing incoherent about the situation in which our prisoner finds herself. What is incoherent in this context is her postulated ability. Similarly, in Tim's case, the incoherence lies not in his situation, a time travel scenario, but rather in his alleged ability to act in such a way that would change the past.

We have to ask, if you claim that Tim (or anyone) has a given power and someone complains that exercising this power leads to an inconsistency, it is simply
not good enough to reply that Tim has the ability alright although he will in fact never exercise it (thereby bringing about the inconsistent consequence). You need to outline what would happen under the circumstances if the ability in question had been exercised, and if the exercising of it would have lead to contradiction, then this is a good reason for concluding that Tim could not do that thing, that is, could not exercise that supposed ability.

The problem is that allowing Tim the ability to kill GF in effect allows him the ability to change the past in so far as the non-death of GF in 1921 is a fact about the past. Lewis is, whether he likes it or not, committed to the principle that if it is possible for Tim to kill GF in 1921 then it is possible for Tim to change the past. What is the status of this ability? In what sense must it then make sense for the past to be changeable?

_Lewis's Version of the 'Cannot' Case_

I have been arguing that Lewis's argument for Tim's ability is not ultimately convincing. This brings us to the 'cannot' case. I want first to endorse the sort of argument Lewis gives for inability, and then go on to provide additional grounds for the inability claim.

(11) It is a fact about the past that GF did not die in 1921.
(12) If it is possible for Tim to kill GF in 1921 then it is possible to render a fact about the past such as not to be a fact about the past.
(13) If it is possible to render a fact about the past not a fact about the past then it is possible to change the past.
(14) It is not possible to change the past.
(15) Therefore, it is not possible for Tim to kill GF in 1921.\(^\text{18}\)

It is clear that this argument, relying upon the impossibility of changing the past, makes a case for the implausibility of an agent's possessing the power to alter the past. The idea that the past is immutable is one of our most firmly held intuitions—and rightly so, most of us would feel. The problem is that the power to alter the past appears to amount to the claim that it is within one's power to bring it about that a given event in the past both did and did not occur in the past. We will look later at another sense of power over the past which seems to avoid this contradictory consequence.

\(^{18}\text{Once again, I have reconstructed Lewis's argument (1986: 75-6).}\)
Other Reasons for Tim's Inability

It is interesting that in considering the case for Tim's inability, Lewis relies solely on the argument from the impossibility of changing the past and does not mention the argument that Tim cannot kill GF because to do so would set up what Horwich calls a *self-defeating causal chain* [hereafter SDCC].\(^\text{19}\) It would be self-defeating in so far as Tim is both biologically and financially dependent upon GF. This is the whole point of making the victim the killer's ancestor. We need consistency conditions sufficient to preclude such SDCCs. Lewis argues that for the story to come out consistently, Tim must somehow or other fail, but '[h]is failure by no means proves that he was not really able to kill Grandfather'.\(^\text{20}\) It is no doubt true that the existence of a Tim able to make the trip at all entails that he did not in fact succeed in killing GF prior to his begetting children. However, William Lane Craig (1991) raises this question: 'But what sense would it make to assert that he could have killed his grandfather, though he did not in fact do so?'.\(^\text{21}\) In fleshing this out, Craig appears to be appealing to the notion of the impossibility of changing the past, yet really he relies on the impossibility of SDCCs. He continues:

For while it makes sense to say, "From the fact that I shall be in Warsaw next year, it follows, not that I cannot commit suicide this year, but only that I do not," it does not seem to make sense to say, "From the fact that I now exist, it follows, not that I cannot kill my grandfather, but only that I do not." In the first case, there are possible worlds in which one commits suicide and fails to be in Warsaw the following year; but in the second case, there are no worlds in which one kills one's ancestor. Killing one's ancestor entails that one's ancestor was not killed. Hence, the act of killing one's ancestor is inherently unintelligible.\(^\text{22}\)

It might be objected that Craig is wrong here: what about the possible worlds where one kills one's ancestor who is then, more or less miraculously, brought back to life in time to procreate? A legitimate worry might be raised of course over whether one could really be said to have *killed* one's ancestor in worlds where he later returns. In any case, Craig is quite clearly appealing to the idea that the act the time traveller is enjoined to perform is incoherent in virtue of causal inconsistency. Here then is yet another reason to claim that Tim cannot do this particular sort of action. (What we

---


\(^{20}\) Lewis (1986: 76).

\(^{21}\) Craig (1991: 141).

\(^{22}\) Craig (1991: 141-2).
have yet to discuss is the precise relationship that obtains between these two forms of inconsistency.)

There are other considerations to take account of. It is clear from what Lewis says about failure not proving inability that he wishes to retain Tim's ability. He asks us to suppose that Tom fails to kill GF's partner for the same reason—whatever it is—that Tim fails to kill GF: 'It does not follow that Tom was unable to. No more does it follow in Tim's case that he was unable to do what he did not succeed in doing'. The implication is clear: Tim (and Tom) have the ability in question. Yet nothing Lewis says here constitutes an argument for this. It seems reasonable to assert that if there is no door, the imprisoned woman cannot possibly leave the room. Now it is true that it does not follow from her failure that she is unable to leave. But this fact is compatible with it being the case that she is in fact unable to leave. If there is no possible exit, an enclosed prisoner cannot leave. This inability does not follow from her failure to do so, but rather from the circumstances of the situation. We might pose the question then: Is Tim's situation like the prisoner's or like Tom's?

Everyone is agreed that Tim will fail. But why? We might want to say that he will fail because (i) success would institute a self-defeating chain along the closed timelike curve since, if GF dies in 1921 neither Tim nor the time machine will ever come into existence (or at least in the manner and at the time in which they did); (ii) it is a fixed and determinate feature of the past (i.e., a matter of historical fact) that GF did not die in 1921: no one can be killed before their death. But it's not just enough to know that he will (must?) fail; the task is to provide a plausible explanatory scenario, devoid of occult thwarting forces, to account for this failure. At first sight, this is a real difficulty. We are left with

... the problem of explaining why it is that I [i.e., the time traveller] cannot fire the gun, or if I can, why it is that I can fire it only in certain directions. Either the gun is not behaving as the normal physical object we take it to be or the notion of voluntary action does not apply in the usual way.

Neither alternative seems palatable. Dwyer replies to this thus: 'There may be countless reasons why the assassination attempt fails but these reasons have nothing to do with guns not behaving as normal physical objects or with voluntary action not

---

23 Lewis (1986: 76).
applying in the usual way'. Lewis agrees: 'he somehow fails' (emphasis added). Why? 'For some commonplace reason. Perhaps some noise distracts him at the last moment, perhaps he misses despite all his target practice, perhaps his nerve fails, perhaps he even feels a pang of unaccustomed mercy'.

What if Tim fails and then decides to try again, and again, and again? Presumably Lewis and Dwyer will say that on each particular occasion a specific, yet entirely commonplace reason will explain the failure each time. We know that all attempts to change the past or institute a SDCC will fail (have failed), but this repeated failure is, to say the least, somewhat surprising and not a little disconcerting. As Horwich observes, '...there is considerable strangeness in this—something ad hoc and unsatisfying about explaining the repeated failures in terms of changes of mind, guns misfiring, and so forth'.

Horwich argues from the implausibility of such continual but fortuitous failure to the conclusion 'that there would not be frequent attempts to instantiate self-defeating causal chains', and he explains this in terms of Gödel's own insistence on the technical impossibility of trips into our local past, while leaving open the possibility of trips into the distant past. However this seems insufficient to preclude the logical possibility of self-defeating causal chains, which is what is at issue, and so I think Horwich has rather missed the point. Continued, invariable failure is much easier to square with the agent's inability to succeed than with his ability, the undoubted truth that (even invariable) failure does not prove inability notwithstanding. It is impossible not to feel that the standard explanation has left something out. Peter J. Riggs, writing very recently, agrees, holding that 'Lewis's account may do for a once only attempt, but is untenable as a general explanation of Tim's continual lack of success if he keeps on trying'.

---

26 Lewis (1976: 150).
27 Horwich (1987: 120-1). Gödel himself relies on the rocket ships failing to attain the requisite velocities (1976: 457). This is somewhat disappointing: we feel that threats to consistency are not to be so easily staved off.
29 Riggs (1997: 52). Riggs virtually rules out a priori the sort of solution I will suggest since he includes among the 'assumptions' which he says 'will place appropriate limits on the domain of the problem and (hopefully) avoid any possible misunderstandings' one which has it that 'human beings do have (at least) limited freedom of action' (1997: 51).
But there is another important consideration: just as GF cannot die in 1921 because he lived on till 1957, so too Tim's failure to kill him can't really be the result of any one or combination of countless possible factors, or any commonplace reason. It must be as a result of one particular factor (or combination of factors): that event, namely, the bee stinging him at just the right moment is no less a fixed and determinate feature of the past than GF's survival. Adherents of the standard view such as Lewis, Fitzgerald, Thom, Dwyer, and Nerlich are all at pains to stress the ordinariness of failure: '...there need be no particular reason for his failure. Something distracts him; he loses interest; the trigger is somehow not pulled. Which accident fends the contradiction off matters not at all.' Yet surely they are required to explain not just that Tim failed for some reason or other, but also why he failed in just the way he did. This leaves us with a dual implausibility to worry about: the possibility of repeated failure and the series of specific reasons for each particular failure.

6.3 The Lewis Resolution

I have been arguing that an examination of the arguments Lewis considers ought to incline us toward the view that Tim cannot kill GF, and that there is no reasonable sense in which he can. Lewis of course relies on a different resolution of the grandfather paradox. A more extensive look at what he does with his two senses of 'can' may help us decide which is the best resolution.

For Lewis there is no contradiction in saying that Tim both can and cannot kill GF: 'They are compatible because "can" is equivocal'. The crucial passage needs to be before us:

Tim's killing Grandfather that day in 1921 is compossible with a fairly rich set of facts: the facts about his rifle, his skill and training, the unobstructed line of fire, the locked door and the absence of any chaperone to defend the past, and so on. Indeed it is compossible with all the facts of the sorts we would ordinarily count as relevant in saying what

---


31 Despite some dissenting voices, it is generally accepted in the literature that a Lewis-type resolution works for this paradox. Typical assumptions made in this standard view are: (i) you can affect but not change the past, (ii) times do not 'come around' twice, and (iii) 'failure' does not entail 'inability'.

32 Lewis (1986: 77).
someone can do ... Relative to these facts, Tim can kill Grandfather. But his killing Grandfather is not compossible with another, more inclusive set of facts. There is the simple fact that Grandfather was not killed. Also there are various other facts about Grandfather’s doings after 1921 and their effects: Grandfather begat Father in 1922 and Father begat Tim in 1949. Relative to these facts, Tim cannot kill Grandfather. He can and he can't, but under different delineations of the relevant facts. You can reasonably choose the narrower delineation, and say that he can; or the wider delineation, and say that he can't. But choose. What you mustn't do is waver, say in the same breath that he can and can't, and then claim that this contradiction proves that time travel is impossible.33

For Lewis the contradiction is merely apparent, generated by an equivocation produced by the term 'can'. With respect to one set of facts Tim can kill GF; with respect to another set he cannot—since this latter set includes inter alia the highly relevant fact that GF did not die in 1921.

This argument of Lewis's is at first glance extremely plausible but it has not gone unchallenged. Joseph Wayne Smith (1985) raises the following objection:

But what about with respect to the total set of facts of the situation? The contradiction arises once more. Lewis may not wish to speak of making considerations of all relevant facts, but nothing prevents us from not conforming to his wishes. A good scientist must surely consider all available facts relating to any matter and not behave as Lewis asks us to behave. But then, this contradiction proves that time travel is impossible. We have been given no reason to believe otherwise.34

Smith is quite right to insist on the importance of taking account of the total situation, i.e., of all relevant facts. But once this is provided it is not the case that the contradiction reappears. The key factor concerns the status of the two sets of facts. I have adopted Smith's talk of 'one' set of facts and 'another' set.35 Similarly, Lewis speaks of 'different delineations of the relevant facts'.36 Such talk, by default, assumes a parity of status for these two sets. We have, it would seem, simply been presented with a choice between two different, but equally acceptable, choices. But they aren't equally serviceable, as Lewis reveals when he refers to the choice as one between 'the narrower delineation ... or the wider delineation'.37 In fact, the latter,

33 Lewis (1986: 77).
34 Smith (1985: 60).
35 Smith (1985: 60).
36 Lewis (1986: 77; emphasis added).
37 Lewis (1986: 77).
'wider' delineation either is, or can reasonably be taken to be 'the total set of facts' demanded by Smith. After all the whole motivation behind Smith's demand is that this (for all intents and purposes) complete set contains the fact of GF's non-death in 1921. Relative to this set (which contains this fact) Tim cannot kill GF—as Lewis admits. Lewis has already provided what Smith claims he is reluctant to offer up.

In this light it is difficult to see what grounds Smith has for his claim that the contradiction will reappear. All we really have here is a reiteration of the belief that it is 'possible' for someone to do something he in fact does not do. I suspect that Smith's feeling that the contradiction re-aris can be explained in terms of his residual intuition that Tim can kill GF. What is operating here is the old commonsense notion that it is unreasonable to suppose that some sort of strange constraint on Tim's freedom to act is involved and that hence Tim ought to be able to kill GF. And when he feeds this into his commitment to a wider delineation of the facts, he seems to generate a contradiction. The appropriate response, however, for someone pushing the virtues of the wider delineation is to give up this residual belief in Tim's ability.

We have seen that Smith's criticism of Lewis is misplaced. However the outcome is not necessarily one of which Lewis would approve. Smith was correct in insisting on a reliance on the notion of ability relative to the widest possible description of relevant facts. (This was one of the lessons of our discussion of the notion of 'within one's power' in the last chapter.) But relative to this description Tim cannot kill GF. What we may well be entitled to conclude in the light of cases like Tim's, and where a time traveller tries to murder his younger self, is not the impossibility of time travel, but rather the impossibility of certain sorts of events happening in the past—along with the appropriate implications for agent freedom. Interestingly, Lewis himself has been criticised as giving too much to the fatalist

---

38 What about the case of Tom, Lewis's other (non-time travelling) assassin? He intends to kill GF's partner. GF's partner is standing conveniently beside GF, just as Tom is crouching no less malignantly beside Tim. 'There's no doubt that Tom can kill his victim' (1986: 75), observes Lewis, and who would disagree? But suppose we have decided that Tim cannot kill GF. The Tom and Tim cases have purposely been developed completely analogously (save for the fact that, unlike Tim, Tom is not a time traveller). But if Tim is unable to achieve his goal, how can Tom, crouching beside him, be able to kill GF's partner? Can Tom kill GF then as well? Couldn't GF fatally step in front of his partner at an unfortunate moment? To see why not, consider again why Tim has to fail. Why couldn't Tim kill GF? Because it was not true that he died on that day. Was it true that GF's partner died on that day? If it wasn't, then Tom can't kill him either.
even in talking about ability relative the facts at all. This relativism whereby '[a]ffirming or denying a power depends on secifying the relative facts' does not allow a sufficiently decisive rejection of fatalism for Cargile (1996).\textsuperscript{39} He claims that the fact that I am not going to do something for you does not have to be excluded in order to allow me consistently to say I could do it. My response is just that, if a B-theory is assumed, this is not true. In fact, it is the most relevant fact that would need to be denied.

Lewis is of course aware that giving up agent freedom, even in these bizarre scenarios, seems to entail fatalism. In fact, he goes on to discuss fatalism (as we noted in Chapter 4) in relation to the types of facts it is relevant to take account of in deciding what an agent can and cannot do. Given that Lewis thinks that pointing out the equivocation of 'can' is sufficient to defuse the argument against time travel, why does he feel the need to come down in favour of Tim's ability? The worry seems to be that if you don't assert this ability, you run the risk of fatalism. It is true that we don't ordinarily count future facts, i.e., facts about the future, as relevant to ability claims.\textsuperscript{40} So is the attempt on GF's life (and his survival) a future or a past fact for Tim? It is difficult to see how we could legitimately opt to ignore this fact in Tim's case. The possibility of a SDCC only underscores this. At the very least, we can say that the murder attempt is not unambiguously a future fact for Tim. For one thing, whether or not he remembers, it is an event in his past—or at least his family's past.

There is the further question: granted that we ordinarily choose quite naturally to ignore facts about the future in assessments of ability, are we entitled to do this if we adopt a B-theory of time which sees the future as no less determinate and real than the past? Further complicating matters is the fact that there is no unambiguous sense of what is past and what is future here. The Special Theory of Relativity (STR) has cast doubt on any legitimate notion of an objective, public time in favour of multiple, equally valid times.\textsuperscript{41} The attempt on GF's life is in Tim's past (as well as his future), both chronologically and causally, but it is in GF's future. GF and Tim will make different assessments as to when this event takes place, varying as their ideal proper time clocks show different readings. And neither is simply right: it is not the case that one time determination is objective and correct and the other subjective and illicit. Both are equally valid. Lewis has made use of the notion

\textsuperscript{39} Cargile (1996: 11).
\textsuperscript{40} Taylor would of course disagree that this was justified.
\textsuperscript{41} Not everyone agrees of course that the so-called lessons of STR I refer to possess the philosophical significance I ascribe to them.
of personal time, roughly analogous to the physicist's proper time.\textsuperscript{42} The idea is that normally personal time and external time run together, but in time travel cases the two become separated. Lewis sees the importance of personal time for explicating Tim's story, despite the fact that being able to tell this story either in terms of the personal time ordering as well as the external one makes it difficult to decide which facts to include, and which to exclude, in determining what is relevant to the question of what Tim can do. Lewis fails to adopt a radical enough approach: for him external time remains 'time itself' and personal time 'isn't really time', and he compares intervals in external time to 'distances as the crow flies' and intervals of personal time to 'distances along a winding path', thereby illicitly privileging external over personal time.\textsuperscript{43} But ultimately what matters is the actual reality of events that exist along Tim and GF's timelines that is more significant than whether we call them past, present or future.

\textit{Proper Time}

This proper time is not only central to making sense of time travel scenarios; it is central to STR. It is, I contend, time in its most meaningful sense. It is time as kept by an ideal clock moving along with an observer or object and so not in motion relative to him or it in any way, nor experiencing a different gravitational field. As such it is the only significant and unambiguous time. There is no global time standard against which all clocks may be judged—there is no 'real time', in the ordinary understanding of this term. When observers in relative motion disagree as to the time, neither determination is correct; rather they are complementary descriptions of motions within a spacetime characterised by STR. In STR, it is distances in spacetime that are invariant. Such distances are called spacetime intervals; they are distances from one location in spacetime to another.

\textsuperscript{42} Lewis's distinction seems to take up Hilary Putnam's earlier (1962) distinction between \textit{objective time} and the relativistic notion of \textit{proper time}. Many writers help themselves to similar distinctions. Some prefer to speak of \textit{public time} rather than objective time. Murray Macbeath prefers 'particular time' to 'personal time' because he talks about the histories of inanimate objects as well as persons. He thinks however that 'If time travel does not take place, there is no such thing as particular time, distinct from external time' (1982: 409-10). Nevertheless he admits that 'particular time does have a scientifically respectable cousin in what physicists call "proper time", a time scale that is relative to a specific system' (p.410). I use the term proper time and believe it is conceptually essential irrespective of the status of time travel.

\textsuperscript{43} Lewis (1986: 69-71). (Emphasis added.)
The expression 'proper time' is not a particularly fortunate translation of Einstein's term 'Eigenzeit'. 'Eigen' literally means 'own', which is a far better indicator of the correct sense. The other problem with the usual rendering is that it is very easy to set proper time over against 'time' itself, without a qualifier. This is what Lewis has done in setting personal time over against 'time itself'. This is misleading, since it is proper time that is time in the deepest sense, and consequently, if you want to use the word in some other, less fundamental sense, you should prefix it with a suitable qualifier. The most important thing to remember is that no proper time is privileged: no proper time (which is to say, no time) is more valid than any other. The problem with distinctions like Lewis's personal/external distinction is that it tends to privilege, even if only tacitly, external time, with personal time appearing to be an artificial construction introduced simply in order to clarify an obscure or confused story. It is external time that is the artificial construct, a fact masked by its congruence with residual classical, commonsense intuitions.

It may be suggested that, while we must give up the notion of a strictly objective time, we can retain the next best thing in terms of an intersubjective, shared time comprising a collection of proper times running more or less in phase. This might be thought useful in time travel scenarios since it would allow us to distinguish between the worldline of the time traveller and the worldlines of the Earth and all non-time travellers:

Thus, that 'thick' world line with fuzzy boundaries which is the Earth's world line is itself made up of a vast number of 'thinner' world lines of varying lengths. We could regard the world line of the Earth as something like a very long tapestry which is composed of many thin threads of mixed lengths. I take it that time travel into the past occurs if the world line of one of these smaller bodies, one of the threads that go to make up the tapestry, extends outwards from the fuzzy boundaries of the Earth's world line and bends back to rejoin the latter world line at an earlier time ... Under such circumstances I want to say that the object has travelled backward in time for a certain period of its history.44

On this basis we can (trivially) distinguish personal time from other proper times. But it must always be kept in mind that the fuzzy, thick worldline is no more real or objective than the thin, time traveller's line. We must guard against an unconscious tendency to slide from 'shared' to 'external' to 'objective' to 'real'. The other point to remember here is that we don't need a concept of objective time, of time itself. The original objective/subjective distinction was supposed to ameliorate the absurdity of

44 Dwyer (1978: 19).
supposing that some event (say, Tim's attempted assassination of Grandfather in 1921) is both earlier than and later than another event (say, Tim's departure in his time machine). Larry Dwyer has spoken of an event like the assassination attempt being such that it lies in the time traveller's 'causal future' but also in his 'chronological past'. But surely we can just as legitimately speak of the 1921 attempt as lying in Tim's causal past (after all, he would neither have been born nor have been able to construct his time machine without the causal fall-out of that attempt's failure), and in his chronological future (since it lies ahead of him on his proper time worldline). This event in fact lies both in Tim's past and in his future as located along his proper time line—which is just as it should be since, ex hypothesi, we are assuming that Tim's worldline describes a closed timelike loop. What we have is not so much a circle as a loop within linear time (assuming Tim goes on his way after the assassination bid), but locally it can be regarded as a circle which is effectively closed by the light ray linking GF in the street and Tim as he lines him up in his sights from his hiding spot. (It is perhaps easier to grasp this point if we imagine Tim meeting himself rather than GF). But, strictly speaking, Tim's worldline is not closed: it would only be so were he to fuse with his earlier self.

The significance of these observations about the appropriate concept of time for this discussion is that seeing the time traveller's timeline as not quite joined will later make it easier to see the appeal of my claim that Tim's predicament is our own. His spacetime path may deviate in ways that ours do not, but essentially his worldline is governed by the same sort of spatio-temporal extension as ours. The fact that his worldline loops over itself just makes the need for consistency among all the events that lie along that line more vivid, but in fact neither more, nor less, compelling than in non-time travelling situations.

6.4 Counterfactual Revisions

After the discussion of 'fatalist trickery', Lewis discusses one last way in which Tim's case might be thought to differ from Tom's. Recall that Tom was introduced in order to render plausible the claim that Tim can kill GF, even if he ultimately (and unsurprisingly) does not. Lewis argues that the cases of Tim and Tom are similar, except that in Tim's case we are more readily tempted to adopt 'a semi-fatalist mode of speech'. He admits however that there does seem to be a special worry with

45 Dwyer (1978: 23).
46 Lewis (1986: 79).
respect to the counterfactual question (what if Tim had killed GF?) that does not apply to Tom's respective counterfactual question:

If Tim had killed Grandfather, it seems offhand that contradictions would be true. The killing both would and wouldn't have occurred. No Grandfather, no Father; no Father, no Tim; no Tim, no killing. And for good measure: no Grandfather, no family fortune; no fortune, no time machine; no time machine, no killing.47

So is the idea of Tim killing GF impossible after all?

If you suppose Tim to kill Grandfather and hold all the rest of his story fixed, of course you get a contradiction. But likewise if you suppose Tom to kill Grandfather's partner and hold the rest of the story fixed—including the part that told of his failure—you get a contradiction. If you make any counterfactual supposition and hold all else fixed you get a contradiction. The thing to do is rather to make the counterfactual supposition and hold all else as close as fixed as you consistently can. That procedure will yield perfectly consistent answers to the question: what if Tim [had] 48 killed Grandfather? In that case, some of the story I told would not have been true. Perhaps Tim might have been the time-traveling grandson of someone else. Perhaps he might have been the grandson of a man killed in 1921 and miraculously resurrected. Perhaps he might not have been a time traveler at all, but rather someone created out of nothing in 1920 equipped with false memories of a personal past that never was. It is hard to say what is the least revision of Tim's story to make it true that Tim kills Grandfather, but certainly the contradictory story in which the killing both does and doesn't occur is not the least revision. Hence it is false (according to the unrevised story) that if Tim had killed Grandfather then contradictions would have been true.49

Before we can assess Lewis's argument here we need to distinguish two senses of power over the past:50

---

47 Lewis (1986: 79).
48 Both Lewis's original (1976) paper and the reprinted (1986) version have "had not" here, but this is obviously incorrect in this context.
50 A third sense of affecting the past would be: actually exercised power to affect the past. An agent is said to have actually exercised power over the past just in case either (i) some past event E has occurred and E's past occurrence is the result of some free action of the agent's at some later time, or (ii) E has not occurred and E's non-occurrence is the result of the agent's free action at some later time which has prevented E's occurrence. This type of power is consistent with the standard response.
(a) Counterfactual power to affect the past.
An agent is said to have counterfactual power over the past just in case either (a) E has in fact occurred but it is within the agent's power to do some action such that, were she to do it, then E would not have occurred, or (b) E has not occurred but it is within her power to do some action such that, were she to do it, E would have occurred.

(b) Power to change the past.
An agent is said to have the power to change the past just in case either (a) W is the actual world, W includes some event E's occurring at some past time t, and it is within the agent's power to bring it about that W does not include E's occurrence at t, or (b) W is the actual world, W does not include the occurrence of E at t, but it is within the agent's power to bring it about that W does include the occurrence of E at t.

We saw in connection with Lewis's explicit argument for Tim's inability that if sense (b) is intended, absurd consequences ensue. Now we are being asked to imagine the counterfactual (rather than the real) replacement of E by not-E (along with whatever other revisions prove necessary). Such 'counterfactual' changes to the past are not obviously impossible. The key question however is whether they actually help the case for asserting that Tim is able to kill GF.

We are talking about the power either to bring about past events that have not, as a matter of actual fact, occurred, or to prevent some of those that actually have occurred. But this power (a) is a power that necessarily is never exercised, otherwise we cannot distinguish it from power (b). That is to say, if it were exercised it would immediately collapse into type (a). The claim that doing some action A would prevent the occurrence of some past event E conjoined with the claim that A has been done implies that E has been prevented. Just as my copy of Finnegans Wake is possessed but never read, so this power is possessed but never used. What is there to be said for a power that is necessarily unexercised? But does it have to be 'necessarily unexercised'? Is my copy of Finnegans Wake necessarily

We should observe in passing that this sense of power to affect the past involves backward causation and backward causation is very likely possible only if the B-theory of time is true. In order for this notion to make any sense the past must be 'still' be there to receive any causal signals 'addressed' to the past. On the accretionist model, the past is perhaps still there in a real enough sense as part of the world's inventory. But it is difficult to see how presentism could accommodate it.
unread? Surely not! It is just unexercised as a matter of (perhaps invariable) fact. We would then argue that Tim's power to kill GF is not necessarily unexercised, merely never exercised in fact—as many do argue. Yet we have to ask ourselves: What would it be for this unexercised power ever to be exercised? If it were ever on any occasion to be exercised then we would have an instance of the past's having actually been changed. Everyone holds this to be impossible. It is not just that Tim doesn't kill GF, but rather that, were he to do so, this would entail changing the past. This is impossible, and thus it is not possible for Tim to exercise his alleged power to kill GF. If it is not possible for him to exercise it, then it is reasonable to call it a necessarily unexercisable power. And this is like an ability we are unable to use—no ability at all! (True, this doesn't imply that Tim can't have the power in question, only that it is not possible for him to use it.)

In a recent article, Kadri Vihvelin has argued in a similar fashion against the idea that someone has the power to kill her baby self:

We should agree that someone can do something, in the relevant sense, only if it's true that if she tried to do it, she would or at least might succeed. And everyone should agree that if someone would fail to do something, no matter how hard or how many times she tried, then she cannot do it ... If someone wants to defend the claim that Suzy can kill Baby Suzy, despite the fact that all her attempts to do so have failed and will continue to fail, then she has to defend the claim that there is at least one occasion on which it's true that if Suzy had tried (one more time) to kill Baby Suzy, she would, or at least, might have succeeded.51

But, she argues, 'this counterfactual is always false'.52 Vihvelin's analysis of the relevant counterfactuals supports my contention here: it is not just that Tim will always fail—he must always fail, because to succeed would immediately set up a logical contradiction whereby an event both would and would not have occurred, and perhaps also institute a later causal inconsistency. It is because these two possible inconsistencies are not on a par that I believe the former is more fundamental.

52 Vihvelin (1996: 319). She does not opt for fatalism since although she believes there are some things a time traveller cannot do, she does want to maintain the distinction between 'will not' and 'cannot' (even for some of the acts of time travellers: '...there are things which they do not do which they nevertheless can do' (1996: 316). I claim however that all actions of time travellers are relevantly similar to the actions of non-time travellers in virtue of a global temporal realism which entails that any successful (per impossibile) attempt to do other than what one does would lead to contradiction.
However let's consider these possible counterfactual revisions a little more carefully. They are put forward as examples of the sort of revisions that would be needed were we to hold that Tim had killed Grandfather. They are needed because if we want to assert, contra the fatalist, that it makes sense to speak of Tim being able to kill GF, then we have to be able to show how, were it in fact to happen, this would not lead to contradictions. Lewis suggests that (i) Tim might have been the time-travelling grandson of someone else, or (ii) he might has been the grandson of someone resurrected after their death in 1921, or (iii) he might have been created new in 1920 with a full set of false memories. The problem is that if any of these revised scenarios were in fact to be the case, then it would no longer possible to speak of Tim having been able to kill GF. Tim qua the time-travelling son of someone else has hardly demonstrated any ability to have killed his real Grandfather. Lewis can of course still claim that he possesses this ability, but it is hard to see how it can be anything other than sheer assertion, a claim made even weaker now by the admission that he has failed to exhibit this ability despite his best efforts. Tim's demonstrated ability to kill someone he imagines to be GF in no way entitles us to claim that he is able to do what Lewis originally claimed he was able to do, namely, kill GF where GF is his grandfather. If we revise so that GF is Tim's real grandfather and is killed by him but is somehow resurrected, then can we really say that Tim killed him, rather than merely having winged him? You can normally only be said to have killed someone who remains dead! Finally, in the case of the Tim newly minted with false past, it certainly looks like he too, while he may perhaps kill someone, is not killing his ancestor. And it is hard to imagine any sufficient revision that would not somehow or other entail either that GF didn't really die or that he wasn't really Tim's ancestor. Lewis recognises the need to revise in order to avoid contradiction, but he fails to see how the very strategy meant to animate his avoidance of fatalism instead serves to undermine the very claim of ability it is the fatalist's task to deny.

The ability claim is thus further undermined. If it makes sense at all to talk about the ability to perform a given act then it must be possible to consider how the world would have to be were that act in fact performed. If it were exercised, the killing both would and would not have occurred. If we want to hold fast just to the occurrence of the killing then the causal bit comes in—it either can't have been a killing or it can't have been GF.

Consider the locked room case: if you hold that one can leave the room and hold that there is no door, then, Lewis would presumably say, of course you get a
contradiction. So you have to revise so that the door exists and is one through which one can leave. But how does someone's acknowledged ability to exit a room with an door through which you can exit provide any support for the claim that he could exit the original room which lacked such a door? The fatalist says there is nothing you don't do but can; he denies that appealing to things you in fact do provides any support for claims about what you could do but don't.

The interesting thing however is that for the fatalist, we are just like Tim—his predicament is ours. We imagine (as does Lewis) that we are more like Tom. But consider this analogy with Tom that Lewis appeals to. As we saw, Lewis introduces Tom, the 'ordinary' agent, into the scenario. He does this for his own purposes, namely, in order to compare/contrast him with Tim, the 'extraordinary' agent. The moral is supposed to be that the extraordinary agent really has the same powers/abilities as the ordinary one. That is to say, Tim can do what Tom can do (even if in fact he won't). But we can turn this around. We can use the acknowledged difficulties associated with the predicament of our extraordinary agent to subvert any naive acceptance of the powers of ordinary agents like Tom (or us!). That is, it is not that Tim is like Tom; rather the analogy goes the other way—Tom is like Tim!

In order to see how the desire to accommodate notions of free agency, even on the part of agents like Tim, can mislead, I want to consider a version of the basic, allegedly paradoxical time travel situation that does not involve such notions. It will emerge that the natural solution in these cases works precisely because there is no need to allow for the aforementioned notions.

PARADOXES THAT DO NOT INVOLVE HUMAN AGENCY

6.5 'Logically Pernicious Self-inhibitors'

There is a problem lurking in the background of all attempts to elucidate stories concerning time travellers killing their ancestors or earlier selves: it is that such scenarios involve agents putatively operating in possession of free will. Our very strong intuition is that, as free agents, what is to prevent these time travellers from shooting their earlier selves? Such causality violation seems impossible. We have hitherto been examining means of avoiding these problems based on the idea that 'did not' does not entail 'could not have'. However it may well be possible to formulate a more formidable objection to time travel. Perhaps the threat posed by self-defeating causal chains can be made stronger if we leave out considerations of
agency and human abilities. We will now consider the possibility of causality violation brought about by the existence of a form of apparatus laid out along some causality-violating curve in spacetime: what we want is some automated machine that will shoot its former self. The claim here is that time travel seems to entail the possibility of the existence of what Paul Fitzgerald has called a 'Logically Pernicious Self-Inhibitor' [hereafter LPSI]. Fitzgerald uses this expression in the context of tachyon paradoxes, but in so far as a tachyon may be regarded as a backward time-travelling particle it is obviously relevant to the sorts of paradoxical situations we have been discussing.

Figure 6.1: LPSI device featuring tachyons.

Essentially, the paradox is generated as follows: T will send tachyon 1 (which may be regarded as a signal or message) to reach R at 2:00 hrs iff it does not receive a signal (viz., tachyon 2) from R at 1:00 hrs. R will send tachyon 2 at 2:00 iff it receives a signal (i.e., tachyon 1) from T at 2:00. So, T receives tachyon 2 iff R sends it; R sends it iff R receives tachyon 1; R receives tachyon 1 iff T sends it; T sends tachyon 1 iff T does not receive tachyon 2. Therefore, T receives tachyon 2 iff T does not receive it. In short, the exchange of tachyonic messages takes place if and only if it does not take place. This paradox clearly does not depend on human agency. Earman: 'A contradiction is generated by asking whether a certain event occurs; we find that it occurs if and only if it does not occur.'

---


54 Hence the reinterpretation principle is irrelevant here: different observers' differing interpretations of the event means nothing in the face of the contradiction allegedly generated.
It is obvious that this device (see Figure 6.1) will not function as it is supposed to, since the plan involves self-contradiction. But, here again, we need to ask: how do we account for this failure? Paradoxes such as these do not depend on human agency, and John Earman has shown that an analogous LPSI-device may be constructed that relies on Gödelian notions of time travel:

... consider a rocket ship which at some space-time point \( x \) can fire a probe which will travel into the past lobe of the null cone at \( x \). Suppose that the rocket is programmed to fire the probe unless a safety switch is on and that the safety switch is turned on if and only if the 'return' of the probe is detected by a sensing device with which the rocket is equipped. Is the probe fired? We find that the answer is that it is fired if and only if it is not fired, which is a contradiction if standard logic holds.\(^56\)

There are a number of strategies that might be adopted in the face of this sort of paradoxical situation.

**Possible Responses**

We might want to conclude that there are time machines and Earman-type LPSIs but that no one ever uses them (or rather, attempts to produce contradictions using them). While it may not be possible to rule this out, this very convenient outcome is highly counter-intuitive. Or perhaps the apparatus somehow or other always fails to function as intended: e.g., the tachyon reflector \( R \) just fails on any given occasion to send its signal, just as if \( R \) were a human agent who might just always fail to carry out her instructions. Once again this is implausible in the long term.

The reinterpretation principle is a device whereby it is claimed that it is always possible to reinterpret the emission of a tachyonic particle of negative energy at a later time as the absorption of a tachyon of positive energy that was emitted at the earlier time. These judgements as to whether emission or absorption is involved would be made differently by different observers. (See Benford, Book & Newcomb (1970: 263). They point out however that while this might work for the case of a single particle, it is possible to construct scenarios where this strategy fails: 'If Shakespeare types out Hamlet on his tachyon transmitter, Bacon receives the transmission at some earlier time. But no amount of reinterpretation will make Bacon the author of Hamlet. It is Shakespeare, not Bacon, who exercises control over the content of the message.' (1970: 265). In our scenario, which involves a genuinely paradoxical causal contradiction ('the exchange of messages will take place if and only if it does not take place'), this principle is even less applicable.

\(^{55}\) Earman (1972: 234).

\(^{56}\) Earman (1972: 231-2).
Perhaps we can appeal to the existence of parallel universes. The time machine or Earman probe always travels to the past of some world other than the world it starts out in. Tim could kill GF in this type of 'multiverse' because he leaves the universe in which GF was not killed. The problem with this is that, whatever else it may be, strictly speaking time travel to a parallel universe isn't really time travel at all. A consistent, this-worldly resolution is far more desirable. Besides, we are offered no real physical explanation for how this splitting takes place.

Maybe there actually are real contradictions in the fabric of reality. Such contradictions may well be rare, but aren't they at least imaginable? It is perhaps not totally impossible to imagine that the probe both is and is not fired. Might 'The exchange of tachyons takes place' or 'The probe is fired' be types of statements that are both true and false? But this line of response is so extreme (for any but a dialetheist) that we would be well advised to consider all the alternatives first. There are in any case at least two other ways of dealing with the problems presented by Earman's LPSI that I want now to look at.

**Horwich's Defence**

First of all, I wish to consider Paul Horwich's claim that Earman's argument involves a modal fallacy. Then I will look at a revised form of Earman's argument for the non-existence of closed timelike loops that has been put forward by Smith and which he claims avoids any fallacy. It will be my contention that while Smith's argument is valid, it does not serve to establish his desired conclusion.

Let \( q = ' \text{closed timelike loops exist}' \); \( p = ' \text{the rocket, probe, sensor and safety switch, and so forth exist and function properly}' \); and \( r = ' \text{the probe is fired}' \). The problem is that the conjunction of \( p \) and \( q \) has the paradoxical consequence \( (r \& \neg r) \). So we have:

\[
(16) \quad \neg \phi(q \& p)
\]

---

57 This 'many worlds' explanation is described in Deutsch and Lockwood (1994) and Deutsch (1997: Chapter 12).
That is to say, q and p are logically incompossible, or there is no logically possible world in which q and p are both true.\textsuperscript{58} But Earman also asserts that p is (physically) possible:

\[(17) \, \neg p\]

Even if we ignore the fact that Earman is sliding between two sorts of 'possible' here it is clear that from (16) and (17) Earman cannot infer $\neg q$, i.e., the non-existence of closed timelike loops. Horwich's point is that there could exist closed timelike curves in the actual world or in any physically possible world in which these rockets, switches, probes and so on, are possible but never in fact either exist or if they do then never in fact function correctly. The proponent of time travel will then maintain that from the fact that timelike loops exist it does not follow that such rockets, switches, probes, etc., cannot exist or function properly, but only that they do not exist or function properly. We may well feel dissatisfied with this argument, sensing its incompleteness, in the absence of any explanation of how and precisely where the apparent incapability to function properly comes from.

**Smith's Version**

Smith (1985) purports to give a non-modal version of Earman's argument that avoids this objection: \textsuperscript{59}

\[
\begin{align*}
(18) & \quad (p \land q) \rightarrow (r \land \neg r) & P \\
(19) & \quad \neg (r \land \neg r) & P \text{ (Law of Non-contradiction)} \\
(20) & \quad \neg (p \land q) & 18, 19 \text{ MT} \\
(21) & \quad \neg p \lor \neg q & 20 \text{ DL} \\
(22) & \quad p \rightarrow \neg q & P \\
(23) & \quad p & P \\
(24) & \quad \neg q & 22, 23 \text{ MP}
\end{align*}
\]

But, as is obvious from (18), it is $(p \land q)$, the conjunction not only of the assumption that closed timelike curves exist but also of assumptions about the rocket, switch and so forth, that entails the contradiction $(r \land \neg r)$. Earman argues that 'we do have good evidence that in our world, rocket ships can be programmed in

\textsuperscript{58} Horwich (1975: 439-40).
\textsuperscript{59} (Smith 1985: 61-2). He notes that it is possible to establish the stronger conclusion $\neg \phi q$ if we replace 'q' by '$\diamond q$' and 'p' by '$\diamond p$'.
a manner similar to the one envisioned'. Consequently, 'although we cannot exclude closed timelike lines on logical grounds, we do have empirical reasons for believing that they do not exist in our world'. 60 What is this 'good evidence'? For Earman it appears to be the fact that 'prima facie, the experiments in question involve only operations which we know to be possible in our world'. 61

It would indeed be ad hoc and quite unsatisfactory simply to appeal to 'a fortuitous series of accidents' to explain each individual failure of the device in the absence of a plausible general empirical reason why paradox machines can't be constructed, as Paul Fitzgerald observes. 62 These would be analogous to the series of accidents that each prevent Tim from carrying out his assassination bid. But surely in the case of Earman's device it is quite clear that it is the plan itself, that is, the envisaged apparatus, that involves self-contradiction. The individual components of the experimental apparatus may well be possible, and indeed functional, yet their conjunction impossible. Individually the components function but their joint operation fails to function. It is not clear how the possibility of the individual components functioning correctly warrants any conclusion as to their joint functioning. Consider a time traveller like Tim: It is possible that Tim could have been killed by someone in 1990, and it is possible that Tim could have killed someone in 1990, and yet not possible that Tim could have killed his earlier self in 1990. Since the set-up itself is contradictory, there is no need to appeal to any regular occurrence of happy accidents. Needless to say, it might well be the case that while the (rolled-up) conjunction p (of the various assumptions that the switch, the rocket, the sensor and the probe each function properly) is self-contradictory, if, say, the sensor was a dud or was disconnected, then in that case it would be possible for some (or all) of the other components to work.

Smith's argument seeks to impugn q on the basis that p is plausible. But since p is just a set-up such that r occurs iff r does not occur, surely it is p that is responsible for the contradiction in its own right. Q would only suffer as a result of this if it could be shown that p is a necessary consequence of q. But this is not done. In fact, p is simply assumed as premise (23). We are given no argument towards the paradox in the sense that we are given no argument to the effect that, if timelike lines were to exist, the set-up outlined in p could be constructed beyond Earman's comments about the individual performances of such components. The plausibility

---

60 Earman (1972: 232).
attached to the usual or normal operations of switches and rockets is insufficient to warrant the assumption of $p$. As a result the paradox can be resolved by giving up $p$. In that case (22) becomes

$$ (22') \quad q \rightarrow \neg p $$

and (23) becomes

$$ (23') \quad q $$

and the conclusion (24) becomes

$$ (24') \quad \neg p. $$

Therefore, in the absence of more compelling reasons to believe $p$ empirically plausible, I conclude that Smith's argument is question-begging. The reason for this is clear: the fault lies with $(p \rightarrow \neg q)$. This premise simply assumes that the explanation for the contradiction lies with the existence of the closed timelike loops (i.e., CTCs).

**Craig's Version**

In order to see if there is a way around this problem available to the opponent of time travel, I will consider one final version of the opponent's argument, one derived from Craig (1991), and again using 'p', 'q' and 'r' as before:

$$ (25) \quad q \land \Box p \rightarrow (p \rightarrow p \land q) $$

$$ (26) \quad \neg \Box (p \rightarrow p \land q) $$

$$ (27) \quad \neg \Box (q \land p) $$

$$ (28) \quad \Box (\neg q \lor \neg p) $$

$$ (29) \quad p $$

$$ (30) \quad \neg q $$

Craig embarks on a lengthy discussion of the truth of (25), which he sees Horwich as disputing. Craig sees (25) as possessing 'some intuitive plausibility' 63 and maintains that Horwich's denial of it depends on a special resolution of vagueness that allows backtracking counterfactuals (i.e., counterfactuals where the truth of the antecedent

---

63 Craig (1991: 147).
implies some adjustment of the past). For Craig, such special resolutions are only appropriate when there exists some relation of conditionship between the state of affairs described in the antecedent and that described in the consequent. And, he seems to be suggesting, it may well be the case that this relation of conditionship is here lacking:

... the construction and proper functioning of the rocket have no effect upon the structure of space-time. Hence, if the time-like loops exist and the rocket and so forth are possible, then it seems that it would be true that if the rocket were to exist, both the loops and the rocket would exist, which results in a self-inhibiting situation. But since it is impossible that were the rocket to exist and function properly, then both it and the time loops would exist, it follows that it must be impossible for the time-loops to exist and the rocket to be possible. Since the rocket is possible, necessarily the time-loops do not exist.64

'If this defense of [(25)] is correct,' says Craig, 'then time travel is impossible.' I believe this concentrates on the wrong place. There is a more vulnerable point at which to attack the argument. Moreover, it is one which focuses more keenly on the salient feature of the situation, viz., what actually causes the contradiction. It does not matter whether (25) is true or not. Where is the argument for (29)? I do not see why any defender of time travel would accept this premise. Once again, it is the assertion of the possibility of the rocket/probe device that is problematical. As Craig says about this argument, it 'yields the desired conclusion that since p is possible, it must be the case that the time-like loops do not exist'.65 I claim that p is not possible: there can be no such successful rocket and probe since its operation entails a contradiction. To the reply that p is only impossible since it exploits the CTC loops, I say that even if this is true, it is the conjunction of rocket and loops then that is the source of the impossibility, not simply the loops. Look again at (28):

(28) □(¬q v ¬ϕp)

This just says that at least one of these two disjuncts has to be the case, i.e., necessarily, either the loops fail to exist or the rocket device is impossible. But they both could be true: the loops might fail to exist and the device be impossible—just what one might expect if it were the fault of the self-contradictory nature of the device operating along the CTC pathways that entailed the contradiction. What are the possibilities with respect to these disjuncts?

(i) \((-q & \neg q)\)
(ii) \((q & \neg q)\)
(iii) \((\neg q & q)\)

In fact, anything except: (iv) \((q & q)\). Why assume that the device itself is possible? The assumption is of course motivated by the desire to show time travel impossible, but this is entirely different from a justification. We are given no argument to the effect that (iii) describes reality any better than (i) or (ii).

6.6 The Art of Rendering Self-inhibitors Less Pernicious

The proponent of time travel will have no problem with situation (ii); but since, as we have just seen, it may be possible to argue that the set-up described by p is only self-contradictory because it is 'laid out along' a closed timelike curve in the first place (i.e., situation (i)), it will be advisable, in the light of this possibility, to consider another possible resolution, one that has been advanced by L. S. Schulman, deriving from the work of Richard Feynman and John Wheeler in the 1940s. A version of this response has also been endorsed by C.J.S. Clarke in his (1977).\(^{66}\) (This occurs within the context of a discussion where a time traveler shoots himself, but non-fatally precisely because his hand wavers at the critical moment as the result of an old bullet wound!) The intuition behind this line is that in a Gödelian universe with closed timelike curves the laws of physics manifest themselves abnormally in such a manner as precisely 'to frustrate the execution of any wish whose accomplishment would create a logical antinomy'.\(^{67}\) It is of course fortunate that things should turn out this way; the appeal of the Wheeler/Feynman type approach is that it provides a physically acceptable basis for explaining the aforementioned 'generally applicable empirical reasons' that Fitzgerald suggested would be needed. The key idea is that the only legitimate requirement of a description is that it should be logically self-consistent, and that there will always be at least one consistent solution. Nor should we be surprised that, in a choice between self-consistent and non-self-consistent solutions, nature always opts for the former in preference to the latter. In the light of these considerations, our worries concerning the explanation of incapability, voiced above, may come to seem less important.

---

\(^{66}\) See his discussion (1977: 101-4).

\(^{67}\) Clarke (1977: 101).
L.S. Schulman (1971) has made use of a continuity of nature postulate, derived from the work of John Wheeler and Richard Feynman, in an attempt to resolve the sort of tachyon paradoxes I sketched above.\(^{68}\) It is possible to see how the same idea could be applied to Earman's LPSI as well.\(^{69}\) Essentially the idea is that any switch (found either in the sensor plate of the tachyon apparatus or the safety switch in Earman's device) must be able to be in intermediate 'non-committed' positions. That is to say, the switch position is a continuous function of the signal intensity, and *vice versa*. If we graph these two curves they must meet at a point that indicates a consistent solution, i.e., a given place where the switch flutters or hovers between on or off. Gregory Benford has used this idea in his science fiction novel, *Timescape*.\(^{70}\) However Earman finds this solution unsatisfying:

Whatever plausibility there is to Schulman's solution, seems to me to derive from the version of the paradox he works with and more particularly from the concentration on the switch. By the very definition of a causal signal, it is clear that whether or not a signal has arrived at some given spatial location is an either-or thing with nothing in between. The paradox can be built around the question of whether or not the signal has arrived at a given spatial location at a given time; and at present I do not see any plausible continuity postulate which will block this form of the paradox.\(^{71}\)

But this is mere assertion: it is precisely Schulman's point that switches can never be 'either-or' things. Any reaction on the part of the switch mechanism may be sufficient to tempt us to admit some signal arrival. And so we may well agree that

---

\(^{68}\) Schulman (1971: 481-2).

\(^{69}\) And also applied to Clarke's scenario.

\(^{70}\) Benford (1982: 91-6). The following puzzle may make the key idea clearer. Imagine there is a mountain climber called Max. He sets out up a particular mountain at precisely 6 am on Sunday morning. He walks at varying speeds during his ascent, sometimes stopping to enjoy the view or pick a flower, but he never goes backwards and never strays from the well-defined path. He attains the summit at precisely 6 pm. The following morning he sets off down the same path at precisely 6 am, walking again at varying speeds, and so on. Again, he arrives at yesterday's starting point at 6 pm. Both trips have taken the same period of time, although presumably comparable sections of the trip took different times to traverse on each day. The task is to prove that there exists at least one point on the path such that he is at exactly that point at precisely the same time of day on both days. Although initially counter-intuitive, it is not difficult to see that this must be the case if we imagine the two trips to be superimposed on each other. Imagine either a ghost or time-travelling 'Max' coming down as Max goes up. They must meet because their paths are continuous.

\(^{71}\) Earman (1972: 236).
the question as to whether or not the signal has arrived is an either-or thing and yet deny that this entails that the result of this arrival issues in an either-or response on the part of the switch. And in the absence of a description by Earman of precisely how a stronger version of the paradox can be constructed it would appear to be premature to dismiss Schulman's 'hung-up' solution out of hand. In the intervening years, new examples of LPSIs have been advanced, but if anything they seem to support the view that self-consistent solutions can always be found.

Playing Billiards

Kip Thorne (1994) has described a 'billiard ball' version of the grandfather paradox developed by the physicist Joe Polchinski.

---

72 It is also significant to recognise that Schulman realises that '[t]he point of this discussion is that history is a set of world lines essentially frozen into space time' (1971: 483). This block universe view is of course a traditional way of characterising the B-theory perspective. But, more significantly, it renders more plausible the solution Schulman is endorsing, since it is easier to imagine these 'frozen' worldlines seeking out a self-consistent solution. They have to fit together.

73 In a recent work, Earman admits that the grandfather paradox 'does point to a seemingly awkward feature of spacetimes that contain CTCs: local data are constrained in a way that they are not in spacetimes with a more normal causal structure'. Despite this admission, he doesn't think highly of the paradox as such and the way it has usually been considered in the philosophical literature. For example, he observes that 'by itself the paradox is of no help at all in assessing the status of these consistency constraints' (1995: 284) and looks solely to physics in order to gauge accurately their shape, content and extent (1995: 309).
The advantage of Polchinski's case is that, analogously to Earman's LPSI, it avoids any taint of free will. But although it seems to frustrate its own operation, a self-consistent solution is available. In the original version (shown in Figure 6.2a), the ball collides with its earlier self after its trip back in time through the wormhole, thereby preventing it from ever having entered the wormhole, and thus generating a paradox. According to Thorne, two of his students, Echeverria and Klinkhammer, managed to resolve the 'paradox' by showing two possible trajectories (see Figures 6.2b & 6.2c) for the billiard ball(s) to follow, each of which is fully self-consistent and consonant with the laws of physics at every point. Thorne subsequently discovered that it was possible to generate perhaps an infinite number of self-consistent solutions. These solutions are analogous to the solution suggested by Clarke: a weakened ball (i.e., weakened by its earlier collision) directs its earlier self in (via a slighter jolt than described in 6.2a) into the 'corner' of the wormhole pocket, thereby in turn explaining why it came out at precisely that angle from the exit of the wormhole, and so on. (They also echo the stories we mentioned in Chapter 1 of Osmo and the servant with the appointment with death in Samarra, a situation where the unfortunate souls met their deaths precisely because of the actions they undertook in trying to evade their fate.)

---

Figures 6.2b&c: The billiard ball that emerges from the left opening at $t_1$ is of course the same one that entered the right opening at $t_3$. I have assigned times in the diagram as they would be measured by a sufficiently well-placed observer in the region of the wormhole. By its own clock, the ball re-emerges at $t_3$ and collides with its earlier self at $t_4$. This is a self-inhibiting device in that the collision prevents the ball from entering the wormhole - the very thing that enabled the ball to collide with its earlier self in the first place. Resolutions (b) & (c) demonstrate two self-consistent solutions whereby the collision of the ball with itself serves only to bring about the initial trip through the wormhole.

Ironically, the fact that a self-consistent solution is possible is ultimately what vindicates the decision to regard Earman's LPSI device as the source of the alleged contradiction. Whether or not a contradiction appears depends directly on the interpretation given to the system (with the role of the CTCs remaining the same). If the device is interpreted in an either/or mode (such that the ball must either go unimpeded through the wormhole opening or else be deflected from entering), then contradiction results. But if we allow an old, weakened through deflection ball to hit its earlier self in such a way (i.e., with less but still adequate force) that it enters the opening but at precisely the right angle, then we can explain the result and no contradiction arises. The fact that the re-interpreted device is able to dispel the contradiction suggests that it was only a poor earlier interpretation that was the source of the initial anxiety. LPSIs cannot work: the salient question is just why this is so. These 'continuity of nature' solutions are meant to be an explanation for the non-functioning of the apparatus. Nature abhors a contradiction, no less than a vacuum. If other outcomes lead to paradox, it is hardly surprising that one of the self-consistent ones actually occurs. Earman expressed a final reservation in his (1995), based on the work of Tim Maudlin, about the 'continuity of nature'
However Riggs (1997) reports that Maudlin has now conceded that he was incorrect in his assessment that a fixed point, self-consistent solution might not always exist.

But does the possibility of an infinite number of self-consistent solutions compromise the fatalist interpretation I have been advancing? Just as for any ordinary event there are a number of consistent outcomes, only one of which will actually occur, the situation is no different here. The time travel case is simply a vivid way of depicting the types of 'accommodations' everyday events have reached. The need for self-consistency acts as a parameter on permissible outcomes. Within these boundaries, the ontological determination inherent in the B-theory's global temporal realism comes into play, carrying the fatalistic conclusion on its back. As Yourgrau observes, in a world characterised by circular time 'there are really no genuine possibilities .. and hence no free will in the traditional sense in which this implies the possibility of doing otherwise. In a universe of pure fact, containing only actualities, the most severe requirement is one of logical consistency...'  

This controlling requirement of logical consistency reflects the situation of the GF paradox: necessarily GF will not be killed since his death would mean changing the past and everyone agrees that that is logically impossible. But everything fits together if we can always find some outcome that preserves consistency. We would do well however to remember our motivation for resorting to the LPSI scenario in the first place: it was that fortuitous failure seemed problematic, especially in the light of the assumption of free human agency in the Tim/GF scenario. Why should things work out so neatly? We can see now that what seems problematic in Tim-like cases, seems entirely natural in LPSI contexts where we are under no compulsion to take considerations concerning the freedom of actions to do otherwise into account. This might mean just that the two sorts of situation are in fact quite different. Or it might suggest that we could solve our problems with Tim's case simply by dropping our compatibilist urge to accommodate the freedom to do otherwise.

Consider for a moment the case of a time-travelling agent who encounters his earlier self. This supposedly free agent tries to shoot himself. We escape any charge of logical impossibility thus: if he decides not to shoot after all, there is no paradox,

and if decides to shoot, then he merely 'wings' himself, i.e., his former self, his poor aim being attributed to the effects of the flesh wound he thereby inflicts (inflicted) on his earlier self. Yet here again, as in Tim's case, we have lingering worries: what if the bullet had been just a little to the right and he had killed himself? The problem arises because of our lingering feeling that, despite the 'victim' not having been killed, the 'agent' was somehow free to kill him, i.e., possessed to power to kill himself. In the LPSI cases, we gratefully accept the 'hung-up' self-consistent solution, but here we remain dissatisfied. The lesson we ought to acknowledge is that the self-consistent solutions in both cases are the only way it could actually have happened. There is a problem with the notion of freedom, although not with the agent's acts of will per se. It is still possible to incorporate the latter as (partial) causes of the consistent outcome. We go awry only because we wrongly persist in viewing people shooting themselves and machines shooting themselves as different. And just as it is senseless to ask why nature didn't prefer the inconsistent over the consistent solution in the LPSI case, it should be regarded as no less senseless an inquiry in Tim's.

A FINAL PARADOX INVOLVING HUMAN AGENCY

6.7 Nerlich's Problem

Thus far, I have been examining the GF and analogous, non-agency paradoxes associated with time travel. An interesting objection to the possibility of time travel has been advanced by Graham Nerlich. I believe that it can be treated in a similar manner to these earlier problems. Nerlich claims that his problem is resistant to the standard response of Lewis and others to such difficulties. I will show that this is incorrect, but in a way that demonstrates once again the need to go beyond this standard response in order to give a full account.

I alluded at the beginning of this chapter to the 4-d context in which discussion of time travel to the past can only make sense. Before we go on, a little more needs to be said about that.

The Metaphysic of Temporal Parts

It is quite clear that time travel seems to entail the physical bilocation of at least some physical objects (and here I take this to include human beings). Bilocation may well be counter-intuitive, but this is not in itself sufficient reason to dismiss it out of hand. Only 4-dimensional objects can be bilocated, in the sense of the same thing being at two different places at the same time. David Lewis explicitly
develops his account of time travel within the context of the 4-dimensional metaphysic of temporal parts. This view regards persisting objects such as apples and people as properly being 4-dimensional spacetime 'worms' comprising spatio-temporal 'parts' or 'slices' or 'stages'. Some philosophers think that any series of temporal parts can be an object; others that certain conditions, such as specifying a 'unity relation' for the object must be met, the most common candidate for this relation being *spatio-temporal continuity*. The apple-at-t₁ which has the property of being green is not strictly speaking the same thing as the apple-at-t₂ which has the property of being red, except in so far as they are seen as being parts of the one spacetime worm. The key idea embodied in this relation of spatio-temporal continuity is that it makes use of the concept of a spacetime path in order to explicate identity for a thing. Some time travel cases rely on the idea of the spacetime 'worm' folding back on itself. Probably only in this way is it possible even to consider that a time traveller might meet himself. It is just this sort of case that Nerlich describes in our now-familiar Gödelian universe featuring *closed timelike causal curves* [CTCs].

*The scenario*

Nerlich discusses, *inter alia*, some problems associated with the notion of these CTCs. He quite rightly points out their strangeness and shows how the A-theoretic 'myth of temporal passage' and the notion of, say, 1970, occurring twice (in the sense of some person x being in 1970 at two (second-order) times) leads to confusion in such contexts. He notes that when we tell the story of such a journey the description of it becomes repetitive, although there is nothing to suggest that the events themselves do so. If someone meets his earlier self, this event occurs only once, though it may be experienced twice. He also quite rightly points out that the notion of an ontically open future is not required in order to make sense of purposive actions. He believes that an agent's actions in such a case would be no more fatally constrained than they ordinarily are. In all of this Nerlich says nothing that is incompatible with what I believe is the standard response to the accusations of paradox often levelled at those who would uphold the logical possibility of time travel.

However Nerlich feels that there *is* something wrong with such time travel scenarios. Let x be a Gödelian time-traveller. The world line of x is almost, but not quite closed. X is therefore able to communicate with, or even act causally upon himself. Nerlich develops his worry thus:

---

If x-old passes a signal to x-young (talks to him, for example) then there is a closed, time-like causal curve in the example, even if no one person, thing or particle has such a curve as its world line. One rather simple way to show how this casts some kind of shadow across the intelligibility of the story is to suppose that there is a thing (let it be a clock) which has a closed history (all the particles which compose it have closed time-like curves as their world lines) and which x always has with him. To change our tale in a simple way, suppose that x-young enters the time machine and goes on his travels as soon as x-old leaves it and that the clock is handed from x-old to x-young as they pass at the entry. So x always has a clock with him. We can suppose this "clock" to be a very crude, simple mechanism. For example, it might be a stick in which x cuts a notch at every mealtime. Now we can certainly say of x, consistently, that he ages, accrues memories, scars and so forth. Whatever he gets, by way of traces, he can consistently keep. This is because his world line is not quite closed. But somewhere, somehow, the stick must lose notches if somewhere, somehow it gains them. For its world line is closed. We can easily see how it could get notched. But how can it get unnotched? There must be no trace of the notch, not merely no visible trace, if we are not to run foul to the general sort of inconsistency mentioned before...What x-old brings back from his journey must be the stick which x-young departs with.80

Nerlich generalises this point, noting that

The spacetime regions within which the Gödelian trajectory lies cannot be so constituted that, at every point j, a message can be sent from j to j+1 (modulo n) but not vice versa. On the contrary, at some stages it must be possible to send a message from j +1 to j and not vice versa. That is to say, the direction of time, of aging, cannot everywhere follow the direction of increasing j. Why not? Because this would require a quantity (age) to increase around any of the (presumably infinitely many) closed timelike curves which lie in the region of spacetime nearby the almost closed timelike curve of x. This, we saw, is contradictory. Hence, somewhere in x's journey, time runs backwards for x, though he himself ages.' 81

Nerlich believes that this part of x's story takes him 'into realms of experiences the like of which we shall never encounter and which strike us as strange almost to the point of incomprehension, and which can certainly lay no claim to physical

---

80 Nerlich (1994: 242-3). Note that Nerlich's choice of a "stick" is not quite so carelessly picked as his text seems to indicate. To have chosen another form of chronometer, such as a more conventional watch, would have rendered his case less dramatic due to the apparently cyclic rather than linear nature of such measurement.

81 Nerlich (1994: 244-5).
realism'. And Nerlich believes that he has shown this to entail, if not a strict contradiction, then at least an unacceptable consequence, on the basis of Einstein's discussion of signalling in his original response to Gödel's paper:

The strength of the Einstein approach is that accidents [the standard method of dealing with these inconsistency problems in these scenarios (RN)] will not save the story from collapse into the difficulties focussed upon. It is not that \( x \) *does* signal ahead from \( j \) to \( j+1 \), nor even that he tries to. The fact that the history of the watch yields a *closed* curve means that *no* quantity can increase from \( j \) to \( j+1 \) (modulo \( n \)) at *every* point \( j \) on the curve. That the curve is timelike and in an arcane geometrical manifold, spacetime, makes no difference to this very simple argument which applies as much to spatial as to spacetime loops. Somewhere on the closed curve ageing must decrease from some \( j \) to \( j+1 \); the timelike direction of signalling, and all that accompanies it, must reverse. 83

It is Nerlich's contention that such considerations cast a "shadow" across the intelligibility of the time travel story. He suggests that a problem exists for the stick that does not exist for \( x \): 'what sense can we make of the fact that, though the watch \( x \) wears *cannot* always register increasing time, the man who is wearing it can and does?'. 84 The decisive issue here is obviously the idea that the stick's world line is closed, whereas \( x \)'s is not.

### 6.8 Defusing the Problem

What are we to make of Nerlich's argument? Can we just respond by saying that there unfortunately will be this divergence between the flow of events and the flow of \( x \)'s consciousness, that the stick can be regarded as somehow or other getting unnotched? This is strange but then we should expect strange things to happen if time travel occurs. But I think that there is a real problem here. It is not just that strange things have to occur, like sticks getting inexplicably unnotched—and remember, not only would the notches have to disappear, but every molecule would have to find its way back to where it was before—but we also have the problem of explaining intelligibly how the man can always register increasing time, but the watch can't. It isn't just the strangeness of what happens to the watch: it is that the strangeness does not carry over to the time traveller as well. Also, at most this suggestion shows that we don't have a strict contradiction here. But it does appeal to

---


phenomena occurring for which we have no likely physical explanation, and surely an alternative explanation that did not require hitherto unsuspected physical anomalies would be preferable.

First, I should say something about the general motivation behind the argument. The time traveller appears to be unable to do something that he should be able to do. *Prima facie* there is no plausible reason why he cannot do something as simple and straightforward as passing a stick to someone—in this case, himself.\(^{85}\) We feel there is nothing to stop x-young using the stick he has been given as his chronometer on the journey. And if he receives it, then surely he *can* use it. Yet, obviously, he *can't*—because it is already (fully) notched, and we know he in fact used the unnotched one as chronometer.\(^{86}\)

I will proceed by presenting an alternative account of what could be thought to be going on in this situation. Its chief virtue is that it provides a consistent account where Nerlich has suggested there isn't one.

We require one final contextual explication: Nerlich's discussion takes place within the context of the 4-dimensional view of objects in time. Nerlich is clearly sympathetic to the B-theory of time, and while this theory usually goes hand in hand with such a view, it is clearly possible to subscribe to the former without the latter. Nonetheless, I think it is clear that whether or not Nerlich would want to subscribe to the 4-dimensional view, it is at least true that the scenario he is considering here requires it. Earlier in his paper Nerlich defended this scenario against the Gödelian claim that it leads to paradox:

---

85 It might be thought that this can hardly be the 'simple and straight-forward' action I have said it to be. After all, we are talking about causally interacting with one's self, or rather, an earlier version of one's self. But this worry is misplaced. The simple action is the passing of the stick between 'them'. Given that we are assuming that it is possible to have the 'two time travellers' present in order to assess whether this is has paradoxical consequences, we can hardly rule it out at this stage as problematic.

86 It is not essential that the stick be fully notched however. All that is required is that there be a stick that is notched and one that isn't. Presumably young-X used the unnotched stick. That, after all, is how it got to be notched in the first place. The contradiction is generated by conjoining the fact that he used the unnotched one with the plausible suggestion that nothing can conceivably prevent his using the already notched one he has just received should he so desire.
Nothing in this description entitles us to say that x is at 1970 at two times. We can say that x is in two places in 1970. I see no objection in that, however. In sum, on the question of consistency in Gödel's world, we can say this: it is possible that x-young and x-old are side by side in 1970 and that x-old greets x-young, joyfully, as a long lost self.87

It is difficult to see how this talk of 'x-young' and 'x-old' could be interpreted as implying anything other than that what we have here are two different temporal parts, or stages, of x.

I mention this point now because I think that it holds the key to the solution. When Nerlich comes to present his own worry about the consistency of this scenario, he fails to treat all the objects involved in the same manner. That is, Nerlich commits the fallacy of equivocation. He talks in terms of x-young and x-old, that is, of 4-dimensional time-slices of the time traveller, yet also in terms of a persisting stick that undergoes various changes, that is, of a 3-dimensional view of a self-identical object that persists and changes. But there can be no greater problem for the stick than for x. Nerlich imagines that this is not so: he holds that the two cases are not the same, because x's worldline is an almost closed timelike curve whereas the stick's line is a closed timelike curve, a closed loop. That is, the stick has a closed history in the sense that 'all the particles which compose it have closed timelike curves as their worldlines'.88 But this is not how things are with the stick: the notched stick does not coalesce or fuse with the unnotched stick—they both go on their own ways. 'What x-old brings back from his journey must be the stick which x-young departs with'.89 But in one sense this is not the case. To imagine that it is, is to assume a 3-dimensional view of sticks. But if the stick is bilocated in the sense of being in two places at the same time then we require a 4-dimensional view. The stick must be interpreted analogously to the time traveller who carries it. X's spacetime path is almost closed, and hence open; the stick's spacetime path is similarly almost closed, and hence likewise open. This is why the stick can keep any traces that it accumulates (just as is the case with x). This also explains why it is that we can't have materially-occupied, closed timelike paths—for the very reasons that Nerlich adduces: their existence would imply that ageing could increase everywhere along such paths, and (as Nerlich observes) this is contradictory.

So in a pertinent sense the stick is not the same one, in the sense that Nerlich intends by declaring them the same. There are 'two' sticks, or much more accurately, two time-stages of the one 4-dimensional stick. It is the fact of there being two stages that is important. Loose talk about the 'same' stick *simpliciter* merely deflects attention from the relevant facts. Since we have helped ourselves to the idea of x-young and x-old and allowed that they be roughly contiguous and able to interact, or causally signal each other (viz., by passing the 'baton' of the chronometer), must we not also speak then of 's-young' and 's-old', or distinguish s₁ and s₂? Nerlich says that what x-old brings back from his trip must be the stick that x-young departed with—and so it must. (We might as well say that x-old must be the same person as x-young.) But it is the same stick in the sense that both are temporal stages of the one 4-dimensional object (just as x-old and x-young are both temporal stages of the one 4-dimensional time-traveller). This identity, however, is entirely consistent with the stick retaining all traces of its ageing; hence Nerlich's inconsistency fails to arise.

It is of course much simpler if x-old does not give x-young the notched stick. Had he not, the parity of cases between x, the time traveller, and s, some other time-travelling clock, might have been more readily apparent. But what we are trying to do is explain consistently how he might have done so. The stick that x-young takes on board *and uses as a chronometer* simply cannot be the one (in the sense that it is a different temporal part of that stick) that x-old hands him as he disembarks. There is nothing, to be sure, to prevent x-young from accepting and retaining this gift. After all, the (fully notched) stick must go somewhere! But, equally, the *unnotched* stick must have come originally from somewhere.

*Object loops versus causal loops*

In order to sort these problems out, we should make a useful distinction, namely, that between *object loops* and *causal loops*. Consider for a moment another time travel difficulty that often surfaces: David Deustch refers to it as the *knowledge paradox*, and it involves what the possible existence of *closed causal loops*. A 25th century, time-travelling art critic arrives unannounced in the studio of an utterly obscure 20th century artist-hack. She has come to meet the greatest master of late 20th century art! Needless to say, the hack is very surprised, not least to hear of his—as yet unsuspected—future glory. In order to convince the artist of the truth of his claims, the art critic produces a portfolio of reproductions of the master's works. The critic is suddenly called back to the future, and, in her haste, neglects to take the portfolio

---

with her. The artist then spends the rest of his life 'painting' the masterpieces for which he is to become (more or less) justly famous five centuries later. Where did the paintings come from? From the reproductions. But where did they come from? From the 'original' paintings, of course. Some critics find this scenario very disturbing since the paintings seem to have arrived (artistically, at least) out of nowhere. Craig regards it as sufficiently serious to warrant the dismissal of the possibility of time travel.\(^\text{91}\)

Where exactly does the loopiness lie? I believe we must distinguish carefully between object loops and causal loops. There is definitely a closed causal loop: the reproductions serve to explain the existence of the paintings, and the paintings serve to explain the existence of the reproductions. But when we consider the objects themselves, their worldlines are not closed or fused to form a seamless whole. The world lines associated with the constituent molecules comprising the material objects (both reproductions and paintings) are not quite joined; rather they 'lie over' each other in an overlapping fashion. The world lines of the particles comprising the materials out of which both the prints and the paintings were assembled extend beyond the strict confines of the loop. Causal loops may well be explicable in various ways.\(^\text{92}\) Object loops however, I would argue, are pernicious.

How is this relevant to our case? Follow the world lines of the relevant artefacts and we readily see the full history of these artistic objects. Here too, in the case of the stick, we must look beyond the confines of the 'closed' causal loop. The object loop would only, strictly speaking, be closed, were the sticks to fuse. As it is, the loop of the stick's world line more closely resembles a coil of rope with one part lying over the other. I therefore contest Nerlich's claim that 'the history of the watch yields a \textit{closed} curve'.\(^\text{93}\) We are only tempted to the view that a closed curve is involved because the signal (= the stick) passing between x-old and x-young seems

\(^{91}\) Craig (1991: 149-50). Interestingly, he comes to this conclusion after not much more than half a page of discussion, after not much more than half a page of discussion, having earlier spent far more (yet somewhat inconclusive) time on 'grandfather' and LPSI paradoxes (cf. 1991: 139-49).

\(^{92}\) For his part, Lewis is untroubled by causal loops, making a plausible case for their being essentially unproblematic (see his 1986: 74). Generally speaking, one way of making sense of these loops would be to adopt a strategy analogous to one adopted by atheists in the face of the the cosmological argument for the existence of God: once you have explained each item in a unified whole such as the world, it is illegitimate to persist with further requests for an explanation of the whole itself.

\(^{93}\) Nerlich (1994: 245).
to close the curve. But does it? A 4-dimensional view of the stick shows that this object cannot serve to close a curve that needs closing for Nerlich's argument to work. And consider: if the two temporal parts of the time-traveller, x, do not close the loop, \(^94\) then why should the stick, which ought to be similarly treated, do so?

How can we describe what is going on? In 1970, x-old returns, greets the departing x-young and the stick is passed between them. At this time there are three sticks present: 1. the one x-young brings with him; 2. the one x-old hands to x-young (let us assume that x-young puts this one in his pocket); 3. the one x has used on his journey and which he passed to himself and which lies in his pocket during the journey and then stays with x-old as he proceeds along his post-journey world line. So as of 1970 there exists s\(_1\) (in x-young's hand), s\(_2\) (in x-young's pocket) and s\(_3\) (in x-old's possession). How did s\(_3\) come to be in x-old's possession? Why, it's simply the stick (i.e., s\(_2\)) that had been in x-young's pocket all along! When x-old shuffles off along his worldline after the interchange with x-young, it is s\(_3\) that he takes along with him.

Nerlich's description seems reasonable enough so long as we don't look too closely at it. X-old ought to be able, we feel, to pass the stick to x-young. And x-young ought to be able to receive it. There is an identifiable sense in which it is the same stick as the one he is to use as a chronometer during his journey. If it is passed between them, do we thereby have a closed loop in the stick's worldline? (We may well have a closed causal loop involving the stick, but this is not the same thing.) Where does the stick come from? Nerlich suggests that it comes from x-old, that is, it is the stick that got notched. But we should clarify the question before offering an answer. *Which* stick (i.e., which stage of s) is being asked after when we inquire as to the origin of the stick? The stick x-young brings to the time machine obviously comes from somewhere else—for example, he picks it up from the path leading to the meeting with x-old outside the time machine.

All of this will become clearer if we vary the stick idea. Imagine that x notches not a stick but his arm. Upon disembarking, he hacks off his arm and, with a disarming smile, gives this 'clock' to x-young. X-young puts it in his bag. He can hardly use it as a clock—(a) because it's already notched, (b) because he didn't notch a severed arm, but rather an attached arm, and as Nerlich and others insist in such cases, it is crucial to see that these things do not happen twice (let alone differently a

\(^94\) What about a light ray from old-x to young-x serving to close the loop? But in this case there is nothing to worry about—no increasing quantities—so the paradox does not arise.
second time around). X does not go on the trip twice. While in the bag the arm is of course still aging—it is decomposing, etc. That is, in some B-theoretically acceptable manner, its ideal clock ticks away, measuring out its proper time. Obviously x-old can't give this arm to x-young. After all, x-young received a freshly severed arm, while x-old retains the decomposing arm. The arm's worldline is, I think, clearly seen not to be a closed loop.

We might think: This is all well and good; but what about the time machine itself? Why can't x-young use the time machine that x-old gets out of? Presumably he can't, for the following reason: imagine the clock attached to the time machine itself. When x-young boards it reads t₁₀, the '₁₀' denoting the 10 years x-old's trip took. But if this is the case when x-young starts out, the clock will read t₂₀ when he finishes his journey and this contradicts our original answer of t₁₀. The situation is really the same as in the stick case. X-young can accept the gift of the time machine, just as he can the gift of the stick; but he can't travel in it on pain of contradiction. We need not appeal to strange physical goings on involving the time machine or its clock losing time. But we want to complain: Why can't x-young use x-old's time machine? Well, for the same reason that Tim can't kill GF. And if a Lewisian has no problem with the latter, why should the former be a problem. But Nerlich claims that his problem for time travel goes beyond this latter problem (viz., the GF problem with a Lewis-type resolution).

Given then that x-young can't use x-old's time machine (i.e., his stage of it), where does the time machine that x-young uses come from? Perhaps he builds it himself, or buys it, or finds it somewhere abandoned. The important point is that it a fully consistent story will not come out of the time machine somehow merging with its earlier self to form a closed worldline; nor, as I have shown, is it necessary.

Where has Nerlich gone wrong? He has run together the idea of a closed causal loop to the idea of a closed timelike curve, and thus conflated causal loops and closed object loops. It is true that if x-old passes a signal to x-young (say by talking to him), then there exists a closed, timelike causal curve (even if no one person, thing or particle has such a curve for its worldline) and yet we deny that an object (like the clock) can have a fully closed worldline.

6.9 Four-dimensional Spacetime

Is there a problem with the apparent trilocation of the stick, which occurs for at least part of the time? (It may not occur: x-young might accept the proffered stick, and
then immediately discard it prior to boarding, although its worldline wanders off somewhere.) The 4-dimensional view meshes well with relativity, in that it encourages us to think in terms of 'a homogeneous and undifferentiated spacetime' by which is meant not just 'mere space plus time'. Spacetime is significant because it provides an invariant interval in STR, and the quality of invariance seems to give us information about the deeper nature of reality, especially when set over against the differing perspectives of observers. The reality of our world is not space and time, but an undifferentiated spacetime. The bilocation of physical objects is perhaps problematic in its own right, despite the fact that it can be fitted into a 4-dimensional context that STR renders plausible. Can we make sense of there being one thing in two places at the same time? Nerlich said that he saw no objection in x being in two places in 1970, and by this he meant that two temporal stages of one and the same person could exist at different places at the same time.

Richard Taylor has argued that you can conceive of one thing existing at two different places at the same time along the analogy of how one thing can exist in one place at two different times. The important thing in the latter case is for the object to exist at all times throughout that temporal interval. Had it come in and out of existence during that period, we would be much less willing to admit that the same object has been at one place at two different times. Thus, it is possible for O to exist in one place at two different times when:

\[
\begin{align*}
O \text{ is at } p_1 \text{ at } t_1 \\
O \text{ is at } p_1 \text{ at } t_2 \text{ (where } t_1 < t_2) \\
O \text{ exists throughout the temporal interval } t_1 - t_2
\end{align*}
\]

If this correctly captures what is involved in an object's being at one place at two different times, then the analog for the case of an object's being at one time in two different places will be:

\[
\begin{align*}
O \text{ is at } t_1 \text{ at } p_1 \\
O \text{ is at } t_1 \text{ at } p_2 \text{ (where } p_1 < p_2) \\
O \text{ exists throughout the spatial interval } p_1 - p_2
\end{align*}
\]

It should be immediately obvious that there are countless things that satisfy these conditions and which are therefore things that are at one time in two different

---

96 Taylor (1968: 383).
places. The Nile, for example, but really practically anything at all that is spatially extended. In the case where an object exists at two different times in the same place this is because it is temporally large enough (extended enough) to span the temporal interval from the one time to the other. Once again, there are numerous things that fit the bill, and they need not be as temporally 'large' as the Pyramids. Similarly, an object can exist at two different places at the same time by being spatially large enough to span the spatial interval from the one place to the other.

Someone may object that in the case of the chair that is here last night, and here again today, it is the entire chair that is present on both occasions; but the entire river is not present at either place. Only a very short stretch of its whole length is present at any given place. But the entire chair isn't present either, at any given time, when the chair is at any given place. What it is that is present of the chair today (or last night) is merely a temporal part of the chair, not its complete temporal extent, and what it is that is present of the river at any given place is not the entire spatial extent of the river, but only limited spatial parts of it. It is by identifying "parts" with "spatial parts" and overlooking "temporal parts" that one falls into the mistaken belief that the chair is wholly present at any particular time. Once we recognise that objects like chairs are extended both in time and in space, and that at any one place there exists only a spatial part of an object, and that at any one time there exists only a temporal part of an object, then we can understand how it is that objects can be both in one place at two different times and at one time in two different places. An object can be in one place at two different times if its temporal extension is such that it spans the temporal distance from the one time to the other; an object can be at one time in two different places if its spatial extension is such that it spans the spatial distance from the one place to the other. And if, in this respect, there is no problem with x-young and x-old being at two places at the same time, why should there be any problem for an object like our stick?

Nevertheless, having said that, even though there should be no conceptual difficulty involved in the notion of the same object being at two different places at the same time, this is a slightly misleading way of speaking. Strictly speaking, the person or object is really in two different places at two different times. A better way to express this is to say that the object is just at two different spacetime locations. A good many of the difficulties in the case we have been considering might have been avoided had we but eschewed all this talk of 'different places at the same time' and 'different times at the same place'. Problems arise through failure to attend sufficiently rigorously to the idea that objects do not have spatial locations simpliciter or temporal locations simpliciter. It is misleading to speak of anything
occupying two different spatial locations at the same time. To the extent that we use such discourse we are talking with the vulgar, as Berkeley would say.

In reality, everything is to be understood as if possessing its own ideal clock to measure proper time—the only time there is. Viewed thus, x-young and x-old (or s₁ and s₂) would exhibit different readings. When they meet and x-old passes the 'clock' to x-young, they will disagree as to the time: for x-young it is 1970, but for x-old it is some latter time, say, 1980, given that he has expended 10 years on the trip. Our tendency to think that there is a real or objective time that they are at, is a metaphysical prejudice left over from pre-relativistic thinking. Proper time is the all-important concept here. We only imagine that we have an object (either x or s) that is in two places at the same time because we remain unconsciously wedded to the metaphysical fiction of some common or objective time. I have of course asserted rather than argued for this view, except in so far as its role in providing a consistent explanation for what otherwise appeared an intractable problem constitutes some measure of indirect support.

I will mention just one further advantage of the concept of proper time: it allows us to re-interpret apparent instances of backward causation in time travel contexts. Lewis claims that '...travel into the past necessarily involves reversed causation'. I am not so sure that backwards causation is necessarily involved in (at least) the Gödelian form of time travel we have been considering. The time traveller who has always moved in a timelike fashion could reasonably be said to act on his causal future. The time traveller pulls certain levers in (say) 1990 and arrives in (say) Greece in 500 BC. What reason is there for saying that the arrival in Greece precedes the pulling of levers? From the time traveller's point of view (the causal agent), we have standard forward causation—an earlier event (lever-pulling) causes a later one (arrival). Even if the time traveller, in anachronistic homage to Dr Johnson, kicks the first person he sees in Greece, this does not necessarily mean that the punch precedes the departure. What this view entails is giving up a single ordering of earlier and later, of past and future in favour of many possible time lines—a form of temporal pluralism. We can retain then the notion that an effect cannot precede its cause in proper time. This would seem to accord with the notion of attaching responsibility to those actions I have done: there would otherwise be a worry about holding a time traveller responsible in 500 BC for an assault he committed in 1990, prior to his departure.

97 Lewis (1986: 73).
6.10 A Two-edged Sword

So Nerlich's scenario can be accommodated by the standard resolution of time travel paradoxes. It should be clear by now however, that in this last scenario, just as is the earlier case of murderous Tim, we are better advised to regard agents as fatalistically constrained in various ways. Fatalistic constraint operates in the world of human action no less than in the world of billiard balls. X-young could accept the stick from his older self only on the proviso that he not use it as his clock. In that sense, he was constrained by the implicit threat of causal and physical inconsistency. But in another sense, just like Tim, he is constrained by even stronger considerations—the ontological facts of the matter. I have sketched merely one way in which paradox can be avoided, but the controlling factor would ultimately be what actually happened in the past. X-young has even less chance of flouting this latter constraint.

None of the arguments advanced against the possibility of time travel show it to be impossible. What they do suggest is that, were it to be possible, it might turn out that certain actions could not be performed. But they could not be performed not for any time travel-specific reasons, but rather for ones arising out of the constitution of the real world. The issues operating in these time travel scenarios are really just our everyday ones thrown into sharp relief. Tim's predicament is ours—at every moment of our lives. The additional conceptual risk in the GF scenario (viz., that involving the risk of causal inconsistency) should not go unmentioned, but even without it the important point remains: there is nothing we don't do but can. The intended moral of the Tom character Lewis's story is that the extraordinary agent (Tim) really has the same powers/abilities as the ordinary one (Tom). That is to say, Tim can do what Tom can do (even if in fact he won't). But this sword cuts both ways. We can use the acknowledged difficulties associated with the predicament of our extraordinary agent to subvert any naive view of the powers of ordinary agents like Tom (or us!). That is, it is not that Tim is like Tom; rather the analogy goes the other way—Tom is like Tim! This is demonstrated by the fact that Tom can't kill his grandfather, or Tim's, or anybody if they didn't in fact die at that time. Hence Riggs is wholly incorrect when he claims that

The other would-be assassin in Lewis's story, Tom, is a free agent of this past era. This granted, there is a sense in which Tom can kill Tim's grandfather. If Tom acts freely then, ceteris paribus, he will be able to kill anyone he chooses (including Tim's grandfather) if he is so inclined and sufficiently persistent ... Since Tim does exist it follows that either
Tom did not try to kill Tim's grandfather or that Tom failed even though he could have succeeded. 98

I think enough has been said by now to see why this should not be granted.

CHAPTER 7

BACKWARD CAUSATION & THE FIXITY OF THE PAST

No philosopher, as far as I am aware, has studied the influence of the nineteenth century upon the sixteenth... Does destiny, which may seize upon our existence, and for its own purposes bear us far into the future, never carry us back into the past?

Edward Page Mitchell, 'The Clock That Went Backward' (1881)

In the fifth chapter of that treatise [De omnipotentia], Pier Damiani asserts—against Aristotle and against Fredegarius de Tours—that it is within God's power to make what once was something that once was into something that has never been. Reading those old theological discussions, I began to understand Don Pedro Damian's tragic story...

And so, in 1946, through the working out of a long, slow-burning passion, Pedro Damian died in the defeat at Masoller, which took place between the winter and spring in 1904.

Jorge Luis Borges, 'The Other Death'

7.1 Introduction

There have been numerous science fiction stories that go something like this: the protagonist tells his colleagues that he intends that afternoon to demonstrate for them the operation of a time machine he has invented. He produces a small brass cube and announces his intention to send it five minutes into the past at precisely three o'clock. Sure enough at five minutes to three the cube abruptly disappears from his palm and appears on the table in front of them. One of his audience ponders whether it might not be an interesting experiment to refrain at three o'clock from sending it back. Would this involve some sort of paradox, since manifestly the cube has entered the past in virtue of its journey? They decide to try this out, and at the appointed time, fail to send it. The cube stays where it was—but the rest of the universe disappears!

In this chapter, I examine a similar but far less apocalyptic version of this science fictional experiment. I will look at one specific argument often advanced against the logical possibility of backward causation. This argument trades on the intuitive everyday view that the past is fixed or settled or already part of the definite furniture of the world in a way in which the future is not. This argument aims to undermine any claim that what is so fixed or settled in existence could be caused by

---

1 Although one might think that it would make more sense for it to be bilocated from 2:55 until (at least) 3:00 rather than disappear from his hand.
what is not so fixed or settled or already in existence, namely, the future. It is thought that after the supposed past effect has occurred, one might somehow prevent the alleged cause from occurring. I will investigate what might possibly ensue when one attempts to perform such sabotaging, or *bilking*, experiments and what might explain these results. My suggestion will be that fatalism has a possible role to play in this explanation.

#### 7.2 Bringing About Events

Is it possible for causes to act backwards in time, and thereby produce prior effects? Is it possible for an agent to act now so as to bring about some event in the past? In order to begin we require some minimum account of *what it is for an agent to bring about an event of a certain kind*? We need at least two things: the *power* to act in a certain way; and the *sufficiency* of such an action for the occurrence of the type of event in question. That is, an agent is able, in circumstances C, to bring about an event E by means of an action of kind A, is just the claim:

1. In circumstances C, action A would be sufficient to ensure event E;
2. In circumstances C, it is in the agent's power to perform an action A or not, as he chooses.

Note that the circumstances C should not themselves be sufficient for E. This is just to stipulate that E be not caused by any earlier or simultaneous event. Without this stipulation, we would not be able to assert plausibly that A brings E about.

This bare account says nothing about whether or not E occurs before or after A. But backwards causation is concerned precisely with those cases where E is held to occur *before* A, and it is these that I want to discuss.

#### 7.3 The Backward Causation Hypothesis

Certain constraints must be put in place if in any putative retrocausal situation the backward causation hypothesis is to be regarded as the most plausible explanation for what is going on in any given situation and indeed they are required for any effective test of this hypothesis. Therefore we insist on the following conditions:

---

2 The account of *bringing about* that follows is taken from Price (1984: 300-1).
Condition (a) There must be a regular conjunction between events of two kinds (hereafter A & E)—a conjunction sufficiently remarkable to allow the ruling out of sheer coincidence as a reasonable explanation for this occurrence.

Condition (b) The circumstances surrounding the paired events must be sufficiently well understood to enable the elimination of the existence of forwardly causal chains from E to A. Specifically, A must possess an independently specifiable cause unconnected with E's occurrence (or non-occurrence). I.e., E cannot be causally prior to A and cannot be (even a partial) cause of A.

Condition (c) As noted above, there must be no independent sufficient causal explanation of E which appeals only to prior events. E is not to be caused by any earlier or simultaneous event. The need for this constraint is obvious: were E to be caused by some other (earlier or simultaneous) cause, there would be no need to invoke A at all. The most A could contribute in that case would be to the overdetermination of E. And surely, for the hypothesis of backward causation to be interesting, it has to imply more than that role alone. (We should also note that it is not necessary, in the course of testing a backward causation hypothesis, to rule out any cause, other than A, which is not earlier and not simultaneous with E. This could only be another later cause, and it would indeed be a Pyrrhic victory for the opponent of backward causation to have his argument against an alleged retrocause succeed by failing to rule out some other, non-A retrocause.)

Condition (d) A & E cannot have a common cause. Were it to be the case that they turn out to be the common effects of some earlier cause, we would once again have no need to suggest that A retrocauses E.

The idea is that we are forced to choose between the hypothesis that E caused A and the hypothesis that E has no cause and that the observed correlation between Es and As is entirely inexplicable. The former hypothesis is, the retrocausalist argues, the more appealing.

A Note on Theories of Time

It is possible to view the whole dispute about backward causation as essentially one about the nature of time—i.e., whether an A-theory or a B-theory is true. The A-theory affirms the passage of time in some sense and the notion of the future as a realm of possibilities and so fully or at least partially unreal; the B-theory denies that time passes and affirms the full and determinate reality of the future, seeing it as just
as real ontologically as the past (or present). In this context it is often claimed that backward causation is only possible at all on the assumption of a B-theory, rather than a A-theory, of time. The idea is that we require passage in order to have any bringing about of past effects and also that (as yet) non-existent future causes would, in any case, be unable (because unavailable) to bring about past effects. I will assume that this is correct, and thus that backward causation requires a B-theoretic view of the future. This raises an interesting point however, since the motivating insight behind bilking attempts would seem to be the notion that one might preclude the occurrence of the alleged cause, after the supposed effect has occurred, but before the appropriate time for the cause. The difficulty is that this approach to talking about time is coloured by with A-theoretic intuitions. I merely suggest that there may be some conceptual tension involved in the process itself. This is precisely the sort of tension that fatalism is well-placed to resolve in so far as it views both past and future events as fixed in exactly the manner suggested by the B-theory.

7.3 The Bilking Experiment

Arguments are sometimes advanced to prove that it must be the case that E can never precede A. It is an argument against the logical possibility of backward causation. Of these, the most popular is the argument based on the so-called 'bilking experiment'. It trades on the suspicion that were we to allow for the possibility of an action A (to be undertaken tomorrow) being the cause of an event E that actually occurred yesterday, then it ought to be possible (in accordance with (2) above) to refrain from carrying out the action in question.

The General Form of the Bilking Argument

The 'bilking experiment' argument goes like this: Suppose an agent claims to be able to bring about an event E by means of a (temporally later) action A. In principle, it should be possible to discover whether E has occurred before the time specified for the performance of A. Why? Because the time for action is later than the time of the alleged effect. The agent is then required to perform A iff E has not occurred. Either the agent will succeed or he will fail.

3 Is this true? It is said that you can't talk of backward causation in an A-theory context since the future cause does not exist in order to bring about the past event that is supposed to be its effect. Yet even on the A-theory it is possible that the alleged future cause will exist at some later time, and this might just be enough. The passage aspect remains problematical however.
Successful Bilking. If he succeeds then either A & ¬E, or E & ¬A obtains. The former outcome is incompatible with (1), namely, our contention that A be sufficient to ensure the occurrence of E. This would show that the alleged causal link did not exist, since we would have been able to bring about an action sufficient for the occurrence of an event that in fact did not occur. The latter outcome (i.e., E & ¬A) conflicts with the requirement invoked in condition (c), namely, that the prevailing circumstances not be themselves sufficient for the occurrence of E. (It might, of course, be explained by E’s having no cause whatsoever, but this is no less fatal to the retrocausalist's claim.)

Failed Bilking. What then of the possibility that he cannot succeed? What if, once E had failed to occur, the agent discovers that, try as he might, he couldn't bring about the alleged cause A? And what if, E having actually occurred, the agent finds, despite his best efforts, that he simply couldn't refrain from A? But if he cannot succeed then this is incompatible with (2), which claims that he ought to be able to bring A about (or not) at will. How then can the agent be justified in claiming to be able to bring about E by means of A? If it is the case that on every occasion, after E has occurred, it is possible for some agent or some thing to intervene so that the putative cause A fails to occur, then it cannot on any occasion be A that is responsible for bringing about E. So, backward causation is false even when both A and E occur.

7.5 Dummett's Example

The argument can be made clearer by considering a particular example: Dummett's 'dancing chief' scenario from his 1964 paper on 'Bringing About the Past'. In the mythical tribe he describes, the chief performs a certain dance designed to ensure the bravery of an absent hunting party. We are not to consider so much the counter-intuitive notion of causal action at a distance as the even more surprising fact that he continues to dance even after the hunting has ceased. Past experience has told him that failure to dance has been correlated with cowardice whereas dancing has indeed been highly correlated with bravery. (We can assume a constant conjunction of dancing and bravery to make the claim even sharper.) The conclusion he reaches is that the dancing has a retroactive effect. The chief corresponds to the agent in the argument sketched above, the hunters' bravery to E, their cowardice to ¬E, and the dancing to A.

In order for our chief to embark on a bilking experiment, there is a knowledge requirement: it must be possible to discover whether the hunters had displayed bravery or cowardice during the hunt. I will come back to this later. The chief's mission then is to dance iff cowardice occurred. There are only three possible outcomes:

**Outcome 1.** A & ¬E. This destroys the correlation between bravery and dancing, thereby disconfirming BC.

**Outcome 2.** E & ¬A. This too would seem to rule out backward causation, since the hypothesised later cause's not having occurred is conjoined with the occurrence of the earlier 'effect', which latter must have been caused (if it is not an uncaused event) by some other event.⁵

**Outcome 3.** This is really a pair of outcomes, depending on whether E occurs or not: a. E & A and b. ¬E & ¬A. In the first case the chief finds himself unable to refrain from doing A; in the second, he finds himself unable to perform A. Either outcome results in conflict with (2).

But most of us share the feeling that is embodied in (2): 'Surely it is in my power to perform A or not perform A, as I choose'. If A is supposed to be the (temporally later) cause of E, there appears no good reason why we should not be free with respect to A. We may wish to take any observed inability here, following the occurrence of E, to refrain from performing A to indicate that, rather than A causing E, E is in fact the cause of A. Similarly, in the face of E's non-occurrence, the chief's inability to perform A, try as he may, might lead us to think that we had no good reason to prefer an explanation in terms of backward causation to one in terms of the far more familiar forwards causation. In this case we would come to hold that E's non-occurrence causes A's non-occurrence.

So **Outcome 3**, involving the failure to bilk, seems to undo any evidence we may have had for preferring the backward causation hypothesis. This is to deny constraint (b), so it can hardly help the defender of backward causation. It would, of course, explain the inability of the chief either to dance (or refrain from dancing, as

⁵ Might it be explained by some unspecified even later cause? Presumably a bilking experiment would then have to be mounted in order to discredit this later cause—but how can this be done if its identity is unknown?
the case may be).\textsuperscript{6} As Paul Horwich says '...[the chief] will conclude that his
dancing is not, as he had previously thought, an action wholly under his control'
(p.92). Rather bravery (E) is a necessary causal precondition for the dancing (A).\textsuperscript{7}
It is not therefore that the chief's dancing makes the men have been brave, but rather
that his ability to dance depends on the prior fact of bravery.

So either the correlation between As and Bs either goes away (as in 1 and 2)
or it remains but is properly regarded as evidence for forward causation. \textit{Outcomes 1}
and 2 disconfirm backward causation; \textit{Outcome 3} forces us to conclude that \textit{E causes A}.

\textit{A Preliminary Consideration}

It might be argued that the bilking experiment argument simply fails to disprove
backward causation. Even were it to work as intended the argument can presumably
only ever serve to discredit individual cases of alleged retrocausal influence. A
blanket rejection of backward causation, solely on the basis of the bilking argument,
does not appear warranted. Even if it is true that we are sometimes able to prevent
the purported retrocause after the occurrence of its purported effect, this can in no
way demonstrate the impossibility of backward causation: the most it could do is
preclude the possibility of backward causation being the preferred explanation in
\textit{those} cases where it had been (wrongly, it would now seem) alleged to occur. A
rejoinder to this might be: it is always in principle possible to bilk. Every successful
spoiling instance confirms this. Given therefore the \textit{prima facie} plausibility of the
whole bilking procedure the onus of proof must be on those who would deny that
every alleged instance of backward causation is susceptible to bilking. Even if this is
wrong I think we need a more clearcut response. This can, I believe, be provided.
But first I want to point out a couple of important assumptions required by would-be
bilkers. Rejection of the first delivers a natural way of avoiding the bilking

\textsuperscript{6} Cf. Fales (1990: 134): 'If, alternatively, it is impossible to prevent \textit{B} once \textit{A} has occurred, we must
conclude that the occurrence of \textit{A} has some causal effect upon the conditions which produce \textit{B}, and
thus a causal direction from \textit{B} to \textit{A} is ruled out.' Also Faye (1989: 24): 'If, on the other hand, \textit{B}
cannot be prevented from occurring after \textit{A} has occurred, then \textit{B} cannot be said to cause \textit{A}: on the
contrary, \textit{B} will have to be deemed causally dependent on \textit{A}. [Note that both Fales and Faye speak of
\textit{A} \& \textit{B for my \textit{E} \& \textit{A} respectively.]

\textsuperscript{7} Note that if this is the case, the chief can discover whether or not the hunters have exhibited bravery
simply by trying to dance, and seeing if he can.
objection; rejection of the second is involved in another means of escape. I will argue that both means of avoidance are unsatisfactory.

Two Key Assumptions:

1. The Knowledge Requirement  Dummett observes that the whole argument depends on the assumption that it is in principle possible to know of the occurrence of any past event, independently of knowledge of one's own future actions. It is easy to see why this assumption, which I shall call the knowledge principle, is essential: in its absence the chief might deny that it is possible to discover whether E has occurred or not prior to his attempt to bilk. It is in principle impossible to perform a bilking experiment if you deny this.

According to Dummett, bilking experiments are intended to demonstrate the incompatibility of the chief's twin beliefs (i) that the positive correlation between his dancing and the warriors' bravery is evidence that the dancing causes the bravery and (ii) that the dancing is something in his power to do or not do as he chooses, are in fact incompatible. Dummett notes however that it is possible to retain (i) and (ii) so long as the chief is prepared 'to abandon the belief (iii) that it was ever possible for him to have knowledge, independent of his intention to perform A or not, of whether an event E had occurred'. We don't usually give up (iii) because we have a prejudice in favour of its holding good for every type of event. The assumption that past events are open to our inspection is very deeply embedded.

But in the face of some suitably recalcitrant experience we should not feel forced to retract any one of these claims in preference to the others. Therefore, we cannot say, prior to a decision on this matter, that bringing about the past (i.e., backward causation) is impossible. Dummett canvases these options in order to suggest that it is entirely legitimate to protect (i) at the expense of (iii), and argues that we might reasonably come to suspect that accurate information about certain past events (like E) could not be obtained. Suppose the chief learns from very reliable sources that the hunting party displayed regrettable cowardice. Sincerely trying to bilk, he dances. However it turns out that the report of cowardly behaviour was incorrect. A number of instances like this might be sufficient to convince the chief that it is, as a matter of principle, not possible to gain knowledge concerning the occurrence of any event in the past, independently of knowledge of one's own
future actions. Hugh Mellor believes however that it is always possible in principle to obtain the necessary evidence about the past events in question. He argues this on the basis of the principle that 'real events must have spatiotemporally contiguous effects' and be located 'wherever in spacetime their immediate effects are'. So if there really is no way whatever of perceiving the effect at any time before $t_2$, it cannot have occurred before $t_2$. Whether or not he is right about this—and I am not at all convinced that he is—Dummett's manoeuvre seems a bit too ad hoc for even a determined supporter of backward causation to rely upon.

Or is it? John Wheeler's 'delayed choice' modification of the celebrated double-slit experiment suggests a scenario in which it appears that a choice of measurement made by an observer today can determine retroactively the constitution of reality in the remote past. The quantum principle seems to show that there is a sense in which what the observer will do in the particle's future defines what happens in the past.

We are to imagine, as in Figure 7.1, a large galaxy directly interposed between an observer here on earth and a distant but powerful light source (e.g., a quasar). The gravitational field of the galaxy bends the beams of photons (from the quasar) as they pass nearby. We can ask the question: When we detect a photon from the quasar today (on Earth), which side of the galaxy did it pass by on its way here? The commonsense view would have it that this question was settled (i.e., had a definite answer) billions of years in the past when the photon moved past the galaxy. Wheeler says we can actually influence what we are entitled to say about the past by what we choose to measure today. The observer has to make a decision about where to place the photon detector: either at Position A (the 'interference' position, where the two beams cross) or at Position B (the 'separate paths' position, when they are once again spatially separated). Note that she can exercise this choice for each individual photon.

---

8 Dummett actually speaks in terms of one's own future intentions, but as Price points out, the argument requires the formulation he gives, and which I have adopted (1984: 322 fn. 1).
10 Price (1996) however regards Dummett's strategy as 'by far the strongest defense' (1996: 174). Partly this is because he sees it as being most relevant for contemporary science. But even if does accord well with certain experiments—an example of which I will shortly refer to—I remained unconvinced that it is justified here.
If she chooses option A, interference effects indicating the photon took both paths will be observed; if she exercises option B, this will show that the photon took only one of the two paths around the galaxy. The delayed choice experiment seems to demonstrate that which path the photon took (billions of years ago) is determined by which measurement option we adopt now.

What is the significance of the example? It looks like it's a clear instance of the knowledge principle being flouted. It is in principle impossible to bilk this scenario since it is not possible to have any knowledge of the path taken by the photon in the past independently of our decision as to which measurement option to take up. Thus Dummett's manoeuvre looks less strange and more plausible, yet not compellingly so.

There are nagging problems associated with trying to render this strategy less unattractive via examples like this one. For one thing, quantum weirdness seems to account for the oddity of the experimental result in the delayed choice experiment. This scenario is just an apparently retrocausal version of an already counter-intuitive yet nonetheless physically unavoidable result (viz., the double-slit experiment). Nor

Figure 7.1: Wheeler's 'delayed choice' double slit experiment using a galactic lens.
are we even sure that the scenario is legitimately retrocausal, even speculatively. While many physicists are quite happy to see this set-up as an instance of backward causation, not everyone agrees. Richard Healey says that backward causation isn't involved since the dependence in question is logical, not causal, while Wheeler himself believes that this is not an instance where 'present choice influences past dynamics' but rather the lesson is that 'the past has no existence except as it is recorded in the present'.11 Even if we are prepared to see this example as giving some justification to Dummett's strategy, there is no guarantee that the knowledge principle might not remain intact in non-quantum-affected contexts. Therefore we should look elsewhere for an escape route. I will examine later the possibility of retaining (i) at the expense, not now of (iii), but rather of (ii)—which is to say, by withdrawing (2), the claim that it is within the agent's power, in the circumstances, to perform A or not, as he chooses. Before getting to this, I will consider a commonly held position available to the advocate of backward causation. But before I do, I must consider the second key assumption made by the bilking objection since this new escape route proceeds in part by rejecting it.

2. The Fixity Asymmetry John Mackie says of the bilking argument:

This objection rests on the assumption that past and present events and states of affairs are fixed and settled and unalterable, whereas at least some future ones are still to be fixed, as yet undecided. 12 The argument is that if at any time A is fixed while B is still unfixed, B cannot be causally prior to A, because at this time things might still be so decided that B does not occur.13 [Note: Mackie's A = our E and his B = our A]

The bilking objection undoubtedly rests on this assumption: as Graham Oddie observes, 'It is clear that only by insisting on an asymmetry between past and future fixity will the bilking argument go through'.14 Yet a few lines later he says, 'If it is true that the central assumption of the bilking argument is the fixity of the past then

11 Wheeler (1978: 41). This is a very strong claim. But whether Wheeler's interpretation is the correct one or not will not be crucial in this context. The agreed-upon facts of the experiment are sufficient for our purposes.

12 In an earlier version of this chapter, 'The Direction of Causation' (1966:455) Mackie has 'as yet unsettled'.

13 (1980: 178). Mackie, in considering possible responses to this argument, observes that '[w]e may dismiss the fantastic suggestion that past events may not be fixed but still subject to decision...' (1980: 178).

14Oddie (1990: 81).
the defender of backwards causation will have to deny the fixity of the past. But the antecedent here is false, as the previous sentence indicates: the bilking argument rests on a past fixity/future fixity asymmetry. Yet we hear no more from Oddie about this. His target is past fixity. It is open to us however to refuse to accept this asymmetry, a refusal that might take one of two forms—either asserting the fixity of all events, past, present and future, or asserting the possible unfixity of all events, including some of those in the past. Oddie maintains that in order to keep backward causation in the face of bilking arguments we have to eschew past fixity (at the very least—he argues further that even more must go, that even giving up past fixity is not enough). I shall argue instead for global fixity. Mackie regards the suggestion that past events may not be fixed as 'fantastic', while Oddie thinks that it is no more so than backward causation itself. Be that as it may, the presumption in favour of past fixity is so strong that only the absence of any alternative course would warrant giving it up by a determined defender of backward causation. Besides there is great plausibility to the notion that backward causation is only possible in a B-theoretic context, and no B-theorist can countenance a lack of past fixity. Indeed she would rail against the very idea of thinking of any events as somehow less hard or less determinate features of reality than any others.

What is it for an event to be fixed? A standard view would be, I contend, that the mere occurrence of an event fixes it, if it has not already been fixed by the earlier occurrence of a sufficient cause. And no event is fixed until it, or some preceding sufficient cause of it, has occurred. (Note that on a B-theory view this supposed property is just a relational property; on a A-theory it is a real, non-relational property that is acquired at the time (i.e., date) of an event's (or its sufficient cause's) occurrence.) The asymmetrical view of fixity (like that ascribed to by Mackie) would have it that events that have occurred either before or at t are assumed to be fixed at t, any event which is the causal consequence of fixed events at t is itself fixed at t and no other events are fixed at t. Paraphrasing Mellor, we can define fixity thus:

\[
\text{Fixity } F(e,t) = \text{df. } O(e,t) \lor (\exists e^*)(O(e^*,t) \& S(e^*,e))
\]

where O = 'occurs' and S = 'is a sufficient cause of'. It is clear from this that having occurred is sufficient for being fixed, and so if we allow e to be any actual event occurring at any time (a la the B-theory) then we don't need the second disjunct to

15 Oddie (1990: 81).
fix any event at all. Fixed events then will just be actual events, and, *contra* Oddie, fixity will not be a stronger property than actuality or truth.17 Oddie of course is not disposed to allow future and past events into a definition of fixity:

One transparently defective alternative is to allow the future and the past the same role in the definition. According to this not only are all events which in fact have occurred, fixed, but all events which in fact will occur are also fixed. Obviously, all this does is to fix both future and past, and so rule out the possibility of future events fixing past events.18

It is difficult to see what the problem is here. What does Oddie mean by 'fixing' in the sentence where he speaks of 'the possibility of future events *fixing* past events' (my emphasis)? Causing to exist? Bringing about? Determining? He has already admitted that backward causal links could exist in a world with actual future fixed events. He rejects the compatibilist prejudice surrounding these links, claiming to desire a less attenuated, more interesting form of power over the past, a sense in which this power makes a real difference. Nonetheless he concedes that 'such backwards causal links could exist'.19 Oddie prefers to endorse a principle of *future-to-past efficacy*, which entails that 'there may be a past event the occurrence of which is not yet fixed, but which will be fixed by some future event'.20 You can't have this on the standard definition of fixity (which is why he seeks to construct his own); but there is nothing obviously incoherent with the idea of a past/future symmetrical view of fixity.

In any case I fail to see how the notion of backwards causal efficacy is threatened or precluded somehow by future fixity. Consider two past (and therefore presumably equally fixed) events—say, Socrates' act of drinking hemlock and his subsequent death. We don't usually think that the fixity of both precludes the former from being causally efficacious in bringing about the latter. Why then should we think this in the case of backward causation? One might want to answer: but presumably it was once the case that these acts were not fixed, or that one was at a certain time at which the other was not. But does this miss the point? On the B-theoretical view—and it is presumably only possible to talk about backward causation at all in a B-theory context—those past events were never such that they were not fixed. Socrates' worldline exists (tenselessly) in spacetime, and these two

---

18 Oddie (1990: 84-5).
19 Oddie (1990: 83).
20 Oddie (1990: 82).
events on it are equally real. As Larry Dwyer urges, we should resist the temptation, to regard some (past) events as being somehow 'harder' facts than other (future but no less actual) ones.21

I will now examine the most standard argument mounted by those who are concerned to defend the possibility of backward causation in the face of the bilking argument. I call it the 'compatibilist defence' for simplicity, but it should not be confused with any other doctrines that go by that name (although it will be clear that it is similarly motivated to those views that assert, for example, compatibility of freedom with either causal determinism or theological fatalism). This will then lead into a consideration of the fatalist interpretation of this situation.

7.6 The Compatibilist Defence

The principal attraction of this approach, championed by Craig (1991), Faye (1989) and Anglin (1981), is that it tries to preserve the intuitions about freedom enshrined in (2) along with an explanation of the failure of the chief's attempts to bilk that does not, in the process, throw over the retrocausal account in favour of a more orthodox one. It recognises that we are not in fact compelled to see the bilking failure as evidence for forward causation, thus disabling one prong of the bilking objection's attack. It also can accommodate the idea of future fixity.

Many commentators have pointed out the weakness of the reasoning in Outcome 3.22 Faced with a brute constant conjunction of As and Es, and given that we (justifiably) would want to avoid 'a strange sort of preestablished harmony' that is 'empirically bizarre' (to use Fitzgerald's characterizations), we have a choice between the forwardly causal explanation that E causes A and the retro-causal explanation that A (retro)causes E. And why not see the observed inability as just evidence for backward causation? In other words, the option remains open to us to explain the apparent inabilities of the agent without appeal to ordinary forward causation. If backward causation is possible it could well be the case that A is a sufficient condition for E, and since we have stipulated that E not have (in the circumstances) any other cause, then it seems reasonable to hold that the occurrence (or non-occurrence) of E is determined by the occurrence (or non-occurrence) of A. As Faye puts it (using A and B for our E and A respectively),

---

21 Dwyer (1978: 29).

The paradox to which backward causation allegedly gives rise, and which the bilking argument is supposed to highlight, is in fact generated by our inclination to hold on to the belief that we may intervene as regards B while supposing that the occurrence or non-occurrence of B does not influence the occurrence of A. 23

This belief incorporates the intuition, expressed in (2), concerning what is within one's power, along with the idea that this power is unaffected by the occurrence (say) of the prior 'effect' presumably since it has already occurred. Faye continues:

... my point is this: A occurs at \( t_1 \) only if B occurs at \( t_2 \). Even on the assumption that it depends on our free choice whether B or non-B will occur at \( t_2 \), it still holds that whatever we choose to bring about at \( t_2 \) will indeed occur; and the event, be it B or non-B, which we choose to produce, will determine whether A or non-A occurs at \( t_1 \). The case in which we choose to prevent B is therefore invariably a case in which A does not occur. The reason why we believe ourselves in a position to prevent B after A has occurred is that we conflate, on the one hand, the trivial principle that an event will take place whether it is B or non-B, and on the other our undoubted capacity to prevent B at any time we please. The fact of the matter is that our freedom to choose whether B or non-B is going to occur at \( t_2 \) is compatible with our being unable to prevent B under the given circumstances. It is scarcely controversial that we cannot prevent the event that is going to occur at \( t_2 \) from occurring then: this trivially follows from the tautology that what is going to happen is going to happen. At the same time, the event, whether it be B or non-B, that takes place at \( t_2 \) determines whether or not A occurs at \( t_1 \). We need to appreciate this simple logical point if we are to avoid the paradox that ensues upon believing that we are free to bilk the cause after \( t_1 \) without effecting [sic] the occurrence of A or non-A in the process. 24

The dancing's never being present without having been preceded by the bravery is explained by the fact that the dancing is a sufficient condition for the bravery. As Horwich observes, '[t]his fact is compatible with both the orthodox and unorthodox causal hypotheses and does not favor either one of them', 25 and so this invariable connection cannot be used, except stipulatively, to underwrite a forwardly causal rather than a retrocausal interpretation. The fact that it looks like we can't have \( E \& \neg A \) doesn't mean that we ought to feel compelled to construe the causal connection between \( E \& A \) as forward causation rather than backward causation. This failure of compulsion is a quite significant point. We needn't demonstrate that the backward

---

causation interpretation must be adopted, since the bilking argument tries to show that it can't be sustained, and we have done enough if we show that this need not be so.

In fact we might want to claim that the assumption that one may bilk in effect begs the question against backward causation. Anglin notes that the argument 'fails because it presupposes that the past was whatever it was regardless of what occurs in the present or future and this, precisely, is what is at issue'.

The answer to the above argument is that although one could prevent C, the fact that E has occurred means that one does not. If one was going to prevent C then E would not have happened and for the simple reason that its cause was not going to happen. [Note: his C = our A]

This approach rightly recognizes the possibility of maintaining a retro-causal interpretation in the face of the chief's failure, accepting that even in the situation where the actual future is such that (say) the chief dances and this event is no less fixed than the event of the warriors' bravery, this does not stop us from adhering to the intuition (enshrined in (2)) that the chief could refrain from enacting the cause, thereby preserving agent freedom while removing any hint of conflict with (2).

A Worry?

But one can't help having the feeling that this somehow compromises our ability freely to choose, despite Faye's reassurances about 'our undoubted capacity' to prevent an alleged future retrocause 'at any time we please'. How can our freedom to choose with respect to A (or ¬A) really be compatible with our being unable to prevent 'A under the given circumstances' when there will always be such circumstances, and where in retrocausal contexts these circumstances include the fact about the occurrence of A, i.e., not just effects that have already occurred but causes which are yet to be (and yet which will be) enacted.

The problem is that there is an equivocation in the senses of freedom and power being talked about here. 'Our undoubted capacity to prevent B [i.e., the cause] at any time we please' can only be a generalized sense of power or ability and not the power to perform a particular act under given circumstances. Circumstances will

often serve to prevent (in the sense of circumstances being causally incompatible with the performance of an action) or preclude (in the sense of circumstances being logically incompatible, etc.,) the actual exercise of purported general abilities or capacities or powers (which one may be said to retain while the possibilities for their exercise come and go). It seems to me that the only interesting sense is the power to perform a particular act under the circumstances, since after all, these are the only powers that are ever exercised. Assume that E has occurred, and further that A is the actual future. Faye wants to say that the chief's freedom to choose A or ¬A is compatible with his being unable to prevent A under the given circumstances. Consider a standard definition of freedom of action:

\[
\text{FA } S \text{ is free at } t \text{ with respect to performing } A = df. \text{ It is in } S's \text{ power at } t \text{ to perform } A, \text{ and it is in } S's \text{ power at } t \text{ to refrain from performing } A. \]

This captures, I think, the essence of (2), or at least its intention. But it is clear that this merely unpacks the general sense of being free to act. The relevant sense of being free to act here ought to be:

\[
\text{FA' } S \text{ is free at } t \text{ with respect to performing } A = df. \text{ It is in } S's \text{ power at } t \text{ to perform } A, \text{ or it is in } S's \text{ power at } t \text{ to refrain from performing } A.
\]

Hence we ought to modify (2) accordingly

\[
(2') \text{ In the circumstances } C, \text{ it is in the agent's power to perform } A \text{ or it is in the agent's power to refrain from performing } A, \text{ but not both.}
\]

In doing so, we don't lose any ability to influence the past (or the future) since the chief will either exercise a power to dance or a power to refrain from dancing, thereby having an effect, even as he lacks the power to dance and the power to refrain. Note however the reason for this modification: we are not driven to it because the prior occurrence of E means that the chief cannot refrain from dancing. After all, this may not be a case of real backward causation at all—which would indeed be the truth were it that the actual future contains the event ¬A. Future fixity however entails that this future contains an unavoidable state of affairs, either A or ¬A, as the case may be, and it is this fact that must be reckoned with.

\[29\] This definition is taken from Hasker (1989: 136).
I am suggesting that we may want to avoid the bilking trap by denying that the agent is free to choose with respect to A after the occurrence (or non-occurrence) of E. That is to say, we accept that there is a conflict between backward causation and (2) and reject the latter, replacing it with the more acceptable (2'). Usually at this point, accusations of fatalism start flying around.

7.7 Craig's Accusation of Fatalism

William Craig objects that we are wrong to see any lack of freedom here at all. He sees in this reasoning the typical modal fallacy perpetrated by fatalism:

To conclude that because an effect has occurred one is not free to refrain from retroactively causing it is just another instance of the modally fallacious reasoning that plagues fatalism. 30

And again,

What, then, may be said concerning the standard objection to backward causation, namely, that once an effect is given, one might interfere to prevent the backward cause from occurring? It [sic—if?] it is said that one cannot so interfere, then, it is argued, fatalism results. I think that by now the fallacy in this reasoning has become obvious. From the fact that the effect is given, it follows only that one will not intervene, not that he cannot intervene. 31

In our example Craig would want to argue that if the hunters are brave then it necessarily follows that the cause will obtain, but not that it will obtain necessarily. For the bravery is caused by the dancing; therefore the chief will dance. Craig believes that it is the bilking experiment itself that is conducive to fatalism since 'it implies that if something will be the case it could not be otherwise' and that his preferred response undermines the bilking argument while avoiding fatalism. 32

32 (1991: 109). Note that while Craig himself believes that while the standard objections, like the bilking argument, aimed at demonstrating the logical impossibility of backward causation all fail, he nonetheless holds that there are good metaphysical reasons to doubt the real possibility of backward causation. He believes that backward causation entails a B-theory of time, which theory is false, and that according to the A-theory backward causation is ontologically or metaphysically impossible. See his (1991: 150-4 & 156-7).
Response

But I am not claiming that the inability I am concerned with issues from the failure. Craig is quite right to note that from the fact that the effect is given, it follows only that one will not intervene, not that one cannot (granted for the moment of course that A does cause E).

There can be no denying that from the fact that E occurred, it does not follow that A cannot be performed, merely that it will not be performed; nor can it follow from the fact that A will not be performed that it cannot be performed. The bravery occurs at \( t_1 \) only if the dancing occurs at \( t_2 \), and the fact that the chief does dance does not entail that he could not have refrained from doing so. Thus we can admit that neither the occurrence of the effect nor the performance of the cause entail that the agent could not have refrained — and still maintain that it is true that he could not have refrained, i.e., that he was not free with respect to performing or not performing A, because in the given circumstances, including the (future) fixity of A, he lacked the power. The reasons for maintaining this might be ontological or metaphysical, not logical. Consider the case of an agent asked to leave a room (apparently) containing two doors, both of which are closed. Suppose he chooses Door 1. We might want to argue that from the fact that he left via Door 1 it does not follow that he could not have left via Door 2. And we would be right. (We may even imagine a large number of trials in which he always just happens to leave through Door 1.) Even if it turns out that Door 2 is actually locked and completely impassable, we would still be correct to claim that it does not follow merely from the fact that he left via Door 1 that he could not have left via Door 2. But of course it is also true that he could not, in these circumstances, have left via Door 2! Ontologically, the room is such that he is not able to exit through one of the doors. This fact precludes the actual exercise of a particular ability to leave via Door 2, and it is hard to see how observations about one's purely generalised ability to leave rooms through doors would lessen his plight. He is able to leave through the other however. This fact is sufficient for the truth of the claim that it is in his power to leave through Door 1 or it is in his power to leave through Door 2, but it is not true that it is in his power to leave through either 1 or 2, in the sense that this last implies that both possibilities are open.

Now imagine that the world is ontologically such that the chief dances and the hunters were brave (or even that he invariably dances when the hunters have been brave). These are facts about the world, in particular a fact about an act of dancing at \( t_2 \) (i.e., A), and no agent has the power to render that fact false at some time prior to
$t_2$, or indeed at any time. (In saying that one agent has the power to render it false, I am *not* claiming that it is a necessary truth that $A$ occurs; just that it is unavoidably a fact about the world and hence beyond the power of anyone to change.) Therefore, we can quite acceptably talk about the chief's power to act: he is able to perform $A$, which renders (trivially) true the claim 'it is in his power to perform $A$ or it is in his power to perform not-$A$'; but we can't claim that it is within his power to do either, as he pleases. And it seems to me to be at least possible that the chief's situation is analogous to that of the man in the room.

We have agreed that an agent's failure to do $x$ does not entail that she was unable to do $x$—although it is important to realise that inability would entail failure. Unfortunately Craig and others seem at times to think that when you have shown that *does not* does not entail *cannot*, you have vindicated (libertarian or compatibilist) claims about agent freedom. But seeing that failure does not entail inability does not in itself warrant the ascription of ability to the agent. Consider the standard resolution to the grandfather paradox: as we saw, Lewis claimed that the time traveller's failure to kill his grandfather 'by no means proves that he was not really able to kill Grandfather'. By itself, this tells us nothing about the still open question of ability. Yet there is a temptation to suppose that since failure does not entail inability, therefore the time traveller possesses the ability, but just doesn't use it. Most commentators see fatalism as so inherently objectionable that they seize on any way out of bilking problems that avoids it. But it is only a vulgar sort of fatalism that argues that failure entails inability, giving voice the this sort of claim: 'You can't change the past because you didn't'. A non-vulgar reading of this might go: you didn't change the past (i.e., failure) and this just indicates an inability which is grounded elsewhere. Analogous considerations should then apply in our case. So where is the chief's inability grounded? The fixity of future states of affairs is the culprit. The future event, whatever it may be, is an unavoidable feature of the world: this explains the inability—and hence also the failure.

If this is right, then it seems much more plausible to opt (as noted above) to modify (2) thus,

---

33 Lewis (1986: 76).

34 And consider this: if backward causation is possible, and indeed sometimes does occur, then in any such instances bilking attempts must fail. Maybe this is an oddity insufficiently defused by gestures in the direction of failure not entailing inability.
In the circumstances C, it is in the agent's power to perform A or it is in the agent's power to refrain from performing A, but not both.

We need both disjuncts, even though only one will ever be true because of our lack of epistemic access. This option makes clear the sense in which there is no incompatibility between the chief's failure to bilk successfully and (2)—understood now in terms of (2'). It operates on a narrower interpretation of what it is to have something within one's power to perform. It recognises that in certain circumstances it is quite appropriate to deny an agent the ability to do just as he chooses. Nevertheless, I maintain that this restriction on the agent's freedom does not imply that an agent cannot be justified in claiming to be able to bring about E by means of A. Indeed in the circumstances, the supposition of backward causation, far from being incoherent, suggests that in fact, if A is a fixed feature of the future, the agent cannot avoid doing A. And it is not at all obvious that backward causation can be ruled out just because its defender is willing to accept fatalism in order to protect it from a bilking argument. If you find fatalism completely objectionable, this might count against this type of defence of backward causation; but one might also be perverse enough to see it as an argument in favour of fatalism.

7.8 Backward Causation or Fatalism?

Fatalism thus offers a means of avoiding the bilking argument, but it is a metaphysical or ontological fatalism, built upon considerations about time and the world. It doesn't require us to go to the extreme of rejecting past fixity. Why choose this response rather than the compatibilist one? We are not driven to it in order to rescue the situation from degenerating into a fortuitous but inexplicable harmony of events. The compatibilist alternative can do this: after all it insists on the possibility that A causes E. I said before that this view managed to defend backward causation while also protecting our ordinary intuitions about freedom. But is this latter feature an advantage that can be overrated? Do these intuitions have anything to commend them other than the fact of their being very widely held? Why do we value freedom of action so highly? Is it because we like to think that as agents we can make a difference (not by bringing it about that what won't happen does somehow happen) but in the sense that we can bring things about. But we can retain this part of the intuition. While we may talk loosely of wanting to be able either to bring A about (and therefore E) or not bring A about as we choose, really what we should be concerned with is efficacy. This is evidenced, I believe, by the prevalence of certain loose definitions of fatalism—encountered earlier—as the doctrine that we can never have any effect on the future. A large part of why we value freedom of action so
highly is because we are loath to admit that we might not be in a position to bring
certain things about. But once it is recognised that giving up certain abilities does
not mean completely giving up the ability to shape events, it may well be that such
worries will, if not evaporate, at least lose much of their force.

We have now examined two instances where fatalism can play a role in
explicating otherwise difficult conceptual problems. The third instance, the so-called
'Newcomb's Paradox', is one area where I do not think it is usefully applied in any
specific sense over and above whatever general application it would have if it is in
fact, as I have been arguing, true.
CHAPTER 8
NEWCOMB'S PARADOX

What seest thou else
In the dark backward and abysm of time?

Some say 'If God sees everything before
It happens—and deceived he cannot be—
Then everything must happen, though you swore
The contrary, for he has seen it, He.'

The Tempest, I ii 49-50
Geoffrey Chaucer
Troilus and Criseyde
(trans. Nevill Coghill)

What has come to be known as Newcomb's Paradox, or Newcomb's Problem, is a fascinating problem in its own right, yet it is also pertinent to our topic in at least some of its many guises. In particular it has application to the question of theological fatalism, and although I have hitherto avoided discussion of this variety of fatalism (except briefly in Chapter 1), we shall see that there is something of particular interest for our concerns that is thrown up in considering it. That interest notwithstanding, my final assessment on the relevance of this problem for the issue of fatalism will be mainly negative. This is principally because the Newcomb situation is ultimately too artificial to motivate the application of a fatalist strategy.

8.1 An Outline of the Problem

We are to imagine a being (or demon or super-psychologist) possessed of hitherto stunningly accurate predictive powers. The being in question (the 'Predictor') has never in fact made an incorrect prediction, although the possibility of this one day occurring remains. (We can assign a probability of, say, 0.9 to the Predictor being correct in his predictions.) Understandably, we have tremendous confidence in the predictions of this being. Suppose then there are two boxes, A and B, placed before us. A contains a thousand dollars (hereafter $T); B contains either a million dollars (hereafter $M) or nothing at all. You (the 'Chooser') are confronted with a choice: you may take either the contents of B alone or the contents of both A and B. But there is a catch! If the Predictor predicted (and the prediction and placing of the contents of the boxes are both stipulated to have taken place many years before) that you will take A and B, he leaves B empty, so that your net gain is $T if you actually do pick both boxes, and $0 if you in fact pick just B. But if he predicted that you will take only B, then he puts the $M in B, giving you a gain of $1,001,000 if you
actually pick both boxes, and $M if you pick just B. If he predicts that you will pick randomly, he leaves B empty. What are you to do? That is to say, what is it most reasonable to do?

First, let’s look at the pay-off matrix for a standard Newcomb Problem (in which the Chooser is not sure the Predictor is absolutely incapable of error):

<table>
<thead>
<tr>
<th>Prediction: only B</th>
<th>Prediction: both A &amp; B</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is $M in B.</td>
<td>There is no $M in B.</td>
</tr>
<tr>
<td>Take both.</td>
<td>$M &amp; $T</td>
</tr>
<tr>
<td>Take only B.</td>
<td>$M</td>
</tr>
<tr>
<td></td>
<td>$T</td>
</tr>
<tr>
<td></td>
<td>$0</td>
</tr>
</tbody>
</table>

What makes this such an interesting problem is that there seem to be such good reasons available to both points of view. There seem to be two very plausible arguments that are completely at odds with each other as to which choice you should make. (1) Say you take both boxes: the being will almost certainly have predicted this, and therefore will not have placed the $M in B. But if you take B only, by the same reasoning, it is extremely likely that he will have placed the $M in that box. A knowledgeable third party will bet quite confidently that if you open just B you will win $M and that if you open both boxes you will only get $T. And who would deny the rationality of these bets? You conclude that it makes most sense to choose B alone. (2) But, one wants to object: the $M is either already sitting there in B, or it is not! Presumably, since the placing occurred long ago, whichever situation obtains is already fixed and determinate (in accordance with the general fixed and determinate nature of all past events). There are two possibilities concerning what has happened. If the being has already put the $M in B, and you choose both, then you gain $1,001,000. If he hasn't, then you gain only $T. You don't know what he has done, but either way you get $T more than you would have by taking B alone.

Or would you? Consider again the first ('one-boxing') argument. Suppose that everyone in the past who chose just B got $M. And all those who chose both boxes received only $T. After all, although not infallible, the predictor has always been right in the past. In that case you would be crazy not to choose only B. Indeed imagine you were betting on some other Chooser: it would, on past experience, be reasonable to bet on that Chooser if she one-boxes but not if she two-boxes. And it is no less to reasonable, when your turn comes around, to choose in a manner precisely analogous to the way you bet on her. Then again, reconsider the second
('two-boxing') argument, but with a few modifications on the situation designed to convey the force of this reasoning.

Variation (a): Suppose now that A is transparent so that you can see the $T inside. Having decided to take only B, you are giving up $T which you can see sitting there in the other box. Variation (b): Imagine a sequential variant of the original problem. You are allowed to open B first and then decide whether or not to choose A. You open B, find nothing inside. Surely you would be crazy not to choose A in that case. And if you had already found the $M in B, wouldn't it be just as crazy not to take the extra $T as well? But if on this sequential version you would be crazy not to take both boxes in the end, why wouldn't you be just as crazy not to take them now? Variation (c): Suppose further that B has a transparent side such that its contents can be observed by a well-informed Well-Wisher who wants to see you maximise your gain.1 This Well-Wisher obviously wants you to take both boxes — irrespective of whether B contains the $M or not—and it is perfectly rational to act in accord with such advice. However, the simple truth is that all who have followed the advice of such knowledgeable well-wishers in the past have come up short compared to one-boxers.

8.2 Decision Theory Principles

The problem is that the one-box argument and the two-box argument rely on different principles for deciding how it is rational to act. The former appeals to the idea that one should act so as to Maximise Expected Utility [MEUP]. The latter appeals to the Dominance Principle [DP], the idea that you should perform a given action if it is the case that doing this action results in you being no worse off than doing anything else, and if there is at least one possible outcome in which having done this action benefits you more than any of your other options. That is to say, there are various possible states of the world, and if there exists some action A that is best in one state and at least equal in all the others, then one should choose to do A.2

Assign a probability of 0.9 to the prediction being accurate. It is possible to calculate the expected utility (or benefit to the Chooser) of either choice. The

---

1 See Schlesinger (1974) for the Well-Wisher variation. Incidentally, he believes the only way out is to deny that the game is possible since voluntary choices are 'fundamentally unpredictable' (1974:221).

2 Nozick discussed these two principles of decision theory in his original paper on Newcomb's problem (1969: 118).
expected utility of choosing one box only (namely, B) is \((0.9 \times $M) + (0.1 \times $0) = $900,000\). The expected utility of performing the action of choosing both boxes (A and B) is \((0.1 \times $1,001,000) + (0.9 \times $T) = $101,000\). Thus, according to MEUP, one ought to choose one box. However if we consider the two possible world states determined by the Predictor's options: (i) the Predictor predicts that the Chooser will take B alone, and (ii) the Predictor predicts that the Chooser will take A and B, then in this case clearly the Chooser ought to take both boxes. If state (i) obtains, taking both boxes delivers $1,001,000 as against $M; if state (ii) obtains, taking both delivers $T as against $0. Whichever situation obtains, choosing both will make you $T better off.

8.3 Some Possible Responses

The MEUP and the DP parallel the betting argument and the well-wisher argument we mentioned before and seem to prescribe two conflicting strategies for the Chooser. There are many types of solutions to the impasse that have been offered, among them the following:

1. *One-boxing is the rational strategy.* A standard rule of decision theory to the effect that the DP has to be restricted to those cases where the states of the boxes are probabilistically independent of the actions. But since this problem is one where the states are probabilistically dependent on the actions, the DP, which recommends two-boxing, does not apply.

2. *Two-boxing is the rational strategy.* The above restriction on applying the DP is appropriate only when the probabilistic dependence reflects causal dependence of states on actions. Whenever it does not, which must be the case when the states obtain prior to the actions, the DP can and should be applied. What is crucial then is whether or not the Chooser's decision causally affects the contents of box A. Where such influence or affect exists, the Chooser should maximise expected utility. But in

---

3 E.g., Bar-Hillel and Margalit (1972); Cargile (1975); Levi (1975). See e.g., Cargile: 'The crucial question needing determination is this: is the high probability of the predictor's being right independent of the probability of my choosing C [i.e., one-boxing] or choosing ~C [i.e., two-boxing], or is it not?' (1975: 235).

4 E.g., Nozick (1969); Lewis (1979) & (1981). See e.g., Nozick: "... if the actions or decisions to do the actions do not affect, help to bring about, influence, etc., etc., which state obtains, then whatever the conditional probabilities (so long as they do not indicate an influence), one should perform the dominant action' (1969: 131-2).
Newcomb's problem we have probabilistic but non-causal dependence, so the dominant action is indicated.

(3) *Deny the very possibility of the game.* Both lines of argument are valid, so we have a real paradox. The only way out is to deny that the game is possible. As in the case of Russell's Barber, we have a demonstration that either the alleged predictive power or the being itself possessing such power could not be. There could not be a being able to predict free or voluntary human actions or choices.

(4) *Compromise positions.* Some claim that there is no one preferred or wholly adequate solution since '[e]ach of the conflicting intuitions is, in some contexts justifiable.'

(5) *Redescribe the situation to resolve the conflict between MEUP and DP.* Another way to proceed involves viewing the Predictor not as making predictions *as to what the chooser will choose* but rather as making predictions *that are either correct or incorrect.* The Chooser's two best outcomes will be associated with two different states of nature. In this case neither of the two actions dominates the other. Here there is no dominant strategy, and your best choice will depend on what state of nature obtains and its associated probabilities.

(6) Sometimes an attempt is made to vary the parameters of the original problem to see if consideration of some limit variation can help clarify the issues. In what could be called the 'Fully Inflated Newcomb Problem', $M$ is place in box A instead of the usual $T$. It seems like you have nothing to lose now: you must win at least $M$, and possibly much more. Both the MEUP and DP agree that one ought to two-box in this case. But where is the point of transition between the original $T$ and the $M$ where the MEUP crossed over to a new recommendation? It turns out to be at a value of $800,000. So at less than that sum for box A, the (standard) one-boxer should say, in such partially-inflated cases, that two-boxing, which did make sense only a few dollars ago, no longer is reasonable. That seems odd.

8.4 Explaining the Predictor's Success

There are numerous twists and turns that I have not even suggested. The issues quickly become intractable, and I do not propose to contribute to the proliferation of

---


7 See Brams (1983: 50-2).

8 For 'Fully Inflated Newcomb Problems' see Sobell (1988: 21-2).

9 For example, suppose you are a one-boxer who imagines that you can make it more likely (a good thing surely—after all, one-boxers get rich) that you will get $M$ by one-boxing now. This gives
responses. My concerns lie elsewhere. The interesting thing about this debate is that it is being assumed that the Chooser can do precisely that— choose between alternative options. I am interested in the role (if any) played by fatalistic constraint on one's choices in this context. In order to assess this question, we need to look at the reasons for the Predictor's success, and hence at two issues: (a) the question of whether there is any possibility that some form of backward causation plays a role, and (b) the question as to whether the Predictor is (essentially) infallible or merely, as a matter of fact, never errs.

If it is assumed that the basis for the prediction consists solely in something like particular facts accessible to the Predictor but that lie wholly in the past together with the operation of generally accepted physical and psychological laws, then you might be impressed by the argument that nothing the Chooser can now do can affect the basis for the prediction, and hence nothing the Chooser does now can make any difference with respect to whether or not there is any money in B. Presumably then opening both boxes can't make any difference now, so the Chooser ought to open both.

You might even be puzzled over what reason the Chooser could possibly have for one-boxing. It seems it would have to involve a belief that it is very likely that if she one-boxes she gets $M, and very likely that if she two-boxes she gets $T. Even if she is right about this, you might claim she should still two-box. She doesn't know if the Predictor judged that she was very likely to one-box or two-box; but it doesn't matter. If he judged her very likely to one-box, then she was and is very likely to one-box and he has placed the $M in B. If he judged her very likely to two-box, then she was and is very likely to two-box and he has not placed the $M in B. What the Chooser chooses now cannot affect the likelihood then. Therefore she should two-box. Of course, if she can, in her choice now, affect the prediction then, she should one-box to maximise her gain; or, if the Predictor has foreknowledge of her choice, then she should one-box since the Predictor will see this and place the $M accordingly. If, for either of these reasons, the Predictor is simply reacting like a seismograph needle to whatever she chooses then she should make him jump the way that is best for her. But if the Predictor simply judged that she would one-box and has a probability of 0.9 of being correct, then she should two-box. As Mark Sainsbury says,

some support to your belief that you are in fact a one-boxer. But what really matters is the Predictor's evidence that you are a one-boxer (and you can do nothing now to change that) and his belief that ultimately determines whether you get rich or not. See Swain (1988: 405-6).
... [n]o matter what twists and turns of reasoning you go in for now, you cannot affect what the Predictor has already done. Even if you could make yourself now into a one-boxer, it would not help. What mattered was whether you were a one-boxer, or a person likely to become a one-boxer, at the time when the Predictor made his prediction. You cannot change the past.\(^\text{10}\)

Given that we are talking about genuine prediction and not foreknowledge or the Predictor's placement of money as an effect of your action, then you have no reason to one-box.

\textit{(a) Backward Causation?}

But if it is not case of genuine prediction, then perhaps we begin to see a glimmer of an explanation for the incredibly successful run the Predictor has experienced. If the Chooser's choice itself in some way influences or brings about the Predictor's successful prediction, then in one sense you can do something to help yourself. In the world where such backward influence affects the prediction then we should just choose B since this would cause the Predictor to predict accordingly and thus place the $M in B. Failing to do this would cause him not to place the $M in B. It is only to the extent that backward influence is excluded, as in the original version of the problem, that we remain tempted by the DP. But where such influence exists, the Chooser should always act so as to maximise utility. As Robert Nozick, who first brought the problem to the attention of the philosophical world, says:

\begin{quotation}
If one believes ... that there is backwards causality, that your choice causes the money to be there or not, that it causes him to have made the prediction that he made, then there is no problem. One takes only what is in the second box.\(^\text{11}\)
\end{quotation}

Nor is this the only other possible explanation. Suppose that

\begin{quotation}
... one believes that the way the predictor works is by looking into the future; he, in some sense, sees what you are doing, and hence is no more likely to be wrong about what you do than someone else who is standing there at the time and watching you, and would normally see you, say, open only one box, then there is no problem. You take what is in the second box.\(^\text{12}\)
\end{quotation}

\(^{10}\) Sainsbury (1988: 59-60).

\(^{11}\) Nozick (1969: 134).

\(^{12}\) Nozick (1969: 134).
Here Nozick introduces the possibility that it is some human 'precognizer' who somehow or other reads the future. This in turn raises the possibility that the Predictor could in fact be identified with God.

*(b) God as the infallible Predictor*

Craig notes that it is 'almost irresistible to identify Nozick's "being" with an omniscient God and to construe Newcomb's Paradox as an illustration of the problem of theological fatalism', and adds that Nozick himself came to approve such an identification.\(^{13}\) Would this identification compromise our faith in the ability of the Chooser to choose freely? The answer to this depends in part on whether a God/Predictor so conceived would be thought of as possessing absolute infallibility or merely an unblemished record of never having been wrong about anything. If the latter, then essentially there is no real difference from the original case. Even if we regard this God as somehow perceiving the Chooser making her choice, there would always remain the possibility of error. In that case, we will assume essential infallibility, and hence that it is certain in any instance that the Predictor will get it right. In that event, the pay-off matrix takes on a very interesting form:

<table>
<thead>
<tr>
<th></th>
<th>Prediction: only B</th>
<th>Prediction: both A &amp; B</th>
</tr>
</thead>
<tbody>
<tr>
<td>There is $M in B.</td>
<td>$T</td>
<td>$M</td>
</tr>
<tr>
<td>Take both.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Take only B.</td>
<td>$M</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$T</td>
</tr>
</tbody>
</table>

We clear out two possible outcomes and as Sobell says it is now just a case of ordering how much money you would like from the menu.\(^{14}\) There are simply no possible worlds in which the Predictor gets it wrong and the Chooser ends up with $M + $T. This seems like the ultimate freedom, until you realise that the essential binding together of choice and prediction cuts both ways: there is nothing privileged about the event of the choice, so why continue to regard it as the sufficient condition of the prediction, rather than the other way around.

So if God is taken to be essentially incapable of error, then this begins to look like the sort of evidence that might lead one to doubt their certainty in the Chooser's wholly unconstrained ability to choose freely. As Locke (1978) argues,

\(^{13}\) Craig (1991: 206).

\(^{14}\) Sobell (1988: 3-4 & fn, 2, 7-9).
Once the Predictor has made his prediction then either Outcome B or Outcome C will ... be impossible, though the Chooser does not know which. For once the Predictor has made his prediction, that prediction becomes fixed and unalterable: having made the one prediction, it is no longer possible for him to make the other. So given that the Predictor is absolutely infallible, it is at the time of choosing equally impossible, and in just the same sense, for the Chooser to make any choice other than that predicted. One alternative is now closed: the Chooser cannot choose in any other way. That is, if the Predictor is absolutely infallible in the required sense it is impossible for the Chooser to choose other than as predicted; once the prediction has been made it is impossible for there to be a different prediction; hence at the time of choosing only one Outcome is possible, the one where the Chooser makes that particular choice, whichever it is, which the Predictor has predicted.

It might be thought at first sight that this reasoning is not essentially different from that of the truth fatalist who argues that if it was true a thousand years ago (or true timelessly) that the Chooser would choose to one-box, then two-boxing is no longer an option for her. Nor is it any different from the claim that if God had infallible foreknowledge of her choice then it is similarly constrained by that knowledge. But I believe this may be mistaken. This is a difference in this case: it is that God did something about it in this case. That is to say, there is an empirically real aspect of the world that acts an a concrete indicator of the Chooser's lack of ability to choose other than she will—namely, the money that is either in or not in the box. The difficulty for the theological fatalist has always been motivating the claim that a truth about the future or one of God's beliefs was not a sufficiently hard fact about the past. Perhaps the money can play this needed role.

The response to this by the compatibilist on divine foreknowledge should not surprise us. For example, Craig's reasoning appeals once again to the asymmetric semantic relation between a true proposition and its corresponding state of affairs:

Were I to choose otherwise than I shall, different propositions would have been true than are, and God's knowledge would have been different than it is. Given that God foreknows what I shall choose, it only follows that I shall not choose otherwise, not that I can not. The fact that I cannot actualize worlds in which God's prediction errs is no infringement on my freedom, since all this means is that I am not free to actualize worlds in which I both perform some action $a$ and do not perform $a$. The Newcomb Paradox provides no reason for thinking that from

14. There is $1,000,000 in B_2 because I am going to choose $B_2$

15 Locke (1978: 21).
and

15. Were I going to choose B₁ and B₂, the $1,000,000 would not be in B₂,
it follows that
16. I am not free to choose B₁ and B₂.¹⁶

But as we have seen, the unfreedom comes from that fact alone (if it is the case that Craig will choose to one-box). The money lying there in the box is really no more than a pointer towards another real event, the future event of Craig's choosing to one-box. The fact of that event does not need to be mediated through any additional 'truth', divinely foreknown or not, nor through the existence of any other event in the past, or even the fact of the money in the box. It is interesting that Craig says that the Newcomb situation provides no reason for thinking that his (16) follows from his (14) and (15). That very claim demonstrates that he is assuming an ability to choose freely as a starting point. This assumption is in keeping with earlier assumptions we have encountered, where it was assumed S was able to do X simply because it was acknowledged that any inability did not issue from any failure to perform the said ability. Besides, if we were to regard all events as equally real, why can't we see the event of God's prediction as supplying the truth conditions for the corresponding proposition about Craig's choice? In such a world, it might be the case that the semantic relationship upon which Craig relies so heavily exhibits an anomalous symmetry. In that case we might assert that were God to predict otherwise than he did, it is true that different propositions would have been true than are. But they might not be ones that suggest Craig's ability to choose otherwise.

In any case, if this wrong, then what we have in the infallible Predictor Newcomb problem is nothing that differs from the standard instance of alleged theological fatalism. In that case it would seem far easier to avoid fatalism simply by denying the existence of any such God or Predictor. It seems to me that it will always be more plausible (for anyone not already convinced of the truth of fatalism on other grounds) to admit the non-existence of a Predictor of the requisite type than to accept fatalism on this basis alone.

PART FOUR

CONCLUSION
CHAPTER 9
CONTINGENCY, EFFICACY AND DELIBERATION

As long as any action lay in the future, he reflected, it belonged to freedom and seemed to be accountable to human decisions: as soon as that same action had become past, it looked not merely necessary with the necessity of what is plainly irrevocable, gone, dead, but as though it had always been necessary even before being enacted, and it appeared as though human wills had never affected the result, could never have affected the result, save in the way of making it more certain than ever of coming to pass...

Howard Nemerov
The Homecoming Game

9.1 The Consequences of Fatalism

When Aristotle responded to the fatalist's argument in De Interpretatione 9, the first thing that occurred to him was that fatalism had to be false because it appeared to conflict with the reality of three things—deliberation, the capacity for human action and the existence of real contingency in the world. I have referred earlier to a general reluctance to accept fatalism. This may be due in part to the failure of arguments for it to convince, but I believe it also has to do with certain perceived consequences of the doctrine. These three phenomena mentioned by Aristotle all need to be re-considered in the light of a fatalist perspective, but this does not necessarily entail jettisoning widely acknowledged and indisputable touchstones of our experience. In this last chapter, I want to focus on the phenomenon of human deliberation. But it was far-sighted of Aristotle to yoke the three phenomena of deliberation, efficacy and contingency together, for it will become clear in what follows that they are intimately related.

What will emerge from this discussion is that deliberation and efficacy can readily be retained but that contingency is necessarily modified. I have already argued in the introductory chapter that the truth of fatalism need not in any sense entail the inefficacy of human action. It is still entirely proper to see human agents as acting within whatever boundaries fatalism may impose. We will of course usually be largely ignorant in most contexts involving the future. Given this fact, there is no reason why we can't act just as we always have. As for contingency, if fatalism is sound, then contingency in the Aristotelian sense can only be epistemic at best, in the sense that we will judge something to be either determinately real or not, though we remain ignorant as to which it is. This is not to say that it is any less
important. Indeed the epistemic opacity that shields many events from our awareness is precisely what enables us to contemplate acting on them.

It has been my contention that fatalism follows from essential features of the B-theory and its view of reality. If this is right, then those features will entail the non-existence of contingency in anything other than the broadly epistemic sense referred to above. (Both this consequence, and fatalism itself, issue from the essential facticity of the world, the reality of the future being but one feature of this. Only a thoroughgoing defence of the B-theory could hope to motivate acceptance of the non-existence of metaphysical contingency. Such a defence is beyond the scope of this work.) But what of our apparent experience of deliberation?

9.2 Fatalist Deliberation

Deliberation is not something that it is appropriate to direct towards the resolution of past choices. The feeling that it is not appropriate to deliberate about the past is grounded in the belief that past things are now utterly beyond our control. Robin Small claims that it is not just that we cannot undo what has already been done:

... the truth is that we can neither do nor undo what is already been done. It is not just that we cannot change the past, but that we cannot act on the past. Given this fact, deliberation about the past is pointless, because no decision to bring about or prevent any past event could be effective.\(^1\)

I will return to the question of the pointlessness of deliberation in a later section; for now, I want to concentrate on another claim that is often made about the ability to deliberate. Peter van Inwagen, in the course of pondering whether or not deliberation, which no doubt ultimately produces activity, is in itself a form of activity, makes the following claim:

But all philosophers who have thought about deliberation agree on one point: one cannot deliberate about whether to perform a certain act unless one believes it is possible for one to perform it. (Anyone who doubts that this is indeed the case may find it instructive to imagine that he is in a room with two doors and that he believes one of the doors to be unlocked and the other to be locked and impassable, though he has no idea which is which; let him then attempt to imagine himself deliberating about which door to leave by.)\(^2\)

---

\(^1\) Small (1988: 29).

This is certainly a widely held belief among philosophers. Even Richard Taylor holds that '[o]ne cannot deliberate about his own future act, in case he believes the act in question is already inevitable'. This is, he believes, 'a consequence of the fact that one can deliberate only about what he believes to be within his power to do and to forego, and the very point in calling anything inevitable is to deny that this condition exists'. But here's the rub for the fatalist: it would appear that he will be unable to deliberate, or at least that it will be very difficult for him to give an account of what it is for someone to deliberate. For the fatalist must surely believe, as a result of his B-theory convictions, that only one of the putative alternatives is 'open' and that one is 'closed'. The fatalist believes that of every pair of actions up for deliberation, one is inevitable, and the other impossible, even if he may not know which is which. The question is, Is this enough to ground that which he cannot plausibly deny, namely, that agents display deliberative behaviour?

We may pose two questions. Firstly, in any situation calling for a choice or decision, do we need to believe that both alternatives are open in order to deliberate? And secondly, do there actually have to be alternatives open in order for deliberation to make any sense? Contrary to what many philosophers seem to hold, I do not believe that it is a necessary condition for authentic deliberation that one needs to believe that both options are open in any situation involving choice. With respect to the second question, the fatalist must of course answer in the negative (or provide an appropriate explanation for the unreal appearance of deliberation). But even that admission is insufficient to prevent him exhibiting behaviour that would be as near to deliberation as one could hope for in the circumstances. Deliberation can occur even in a world characterised by non-contingency. (I will return to this point in Section 9.4 below.)

Van Inwagen suggests that it is impossible to deliberate about whether to do something if you don't believe that it is possible to perform it. Now the fatalist certainly believes that it is possible to perform actions. This was the whole point behind our insistence on the efficaciousness of human action, even in the context of the truth of fatalism. But van Inwagen obviously means, as his example indicates, that you have to believe not just that one of a disjunctive pair of actions is possible, but that both possibilities are still open, if deliberation is to be possible.

---

3 Taylor (1964a: 74).
Imagine that I am in van Inwagen's room. I believe that one of the doors is locked and impassable. Whether or not it actually is, I believe that it is. Therefore, according to van Inwagen, I cannot deliberate. But why can't I deliberate in these circumstances? I certainly don't see why my belief should preclude reflection on which door to choose to try in order to leave. Deliberation is an intentional activity. As such, it is not a necessary condition for an authentic instance of deliberation that I succeed. I am trying to reach a decision about which door I shall exit by. I know that at least one door is not impassable. So long as I don't know that it is completely impossible for me to leave the room, I am able to deliberate. I can deliberate over whether to exit via one of the two doors, all the while knowing that if I pick the wrong one I will fail. This is because what I am doing, when I deliberate over which door to take, is deliberating about which exit to try to exit through, and that is entirely consistent with one of the doors being locked and impassable. If to deliberate is not to deliberate about actually exiting, but rather to deliberate about which door to try in order to exit, then it may be entirely to the point to deliberate in the situation where I simply don't know whether or not one of the doors is impassable. After all, there is always an outside chance that any given door is impassable, and most people wouldn't use that as a reason for doubting that they could deliberate in the ordinary scheme of things.

Let's look at the meaning of the word. I think this will offer a clue to why it is not essential that I believe in the real existence of alternatives, as well as a reason why some people mistakenly believe that it is possible to deliberate only over acts where one is equally free to perform or not perform them.

The etymological root of the word is libra meaning 'scales' and hence librare, 'to weigh', and hence deliberare, 'to weigh thoroughly' or 'to consider very carefully'. It does not, contrary to appearances, come from, or have anything to do with, liber meaning 'free'. This is, I think, a source of confusion, since the erroneous echo of 'liberty' in the word has misled some to think that if circumstances are restricted in the way that van Inwagen's example suggests, then deliberation is inconceivable.

Applying this to the case at hand, even if I believe (or indeed know) that one of the exits is blocked, I can still carefully weigh up my options. The task that I set myself may be to diagnose the situation. I may be looking for various subtle clues, either in the room, or manifested in the people around me or with whom I have recently been in contact—much as an expert poker player looks for tell-tale signs hoping to gauge some useful information. I may reflect on previous situations in
which I have pondered such matters. Nor does it matter whether I actually can do any of these things: for example, I may not have been in contact with anyone but have just found myself deposited here. There may be no clues at all that might help me to decide between one door or the other as the most likely candidate. I may never have been in, or imagined I would ever be, in this situation. But I will do what I can—even if that is very little. It is not essential to thinking carefully about a conundrum that I be sure I can solve it. Neither is it necessary that, in order to think carefully about what are my most appropriate actions in the circumstances, I not be constrained by the knowledge that one avenue of escape is closed to me. Ultimately, for all my reflection, I may just choose one and succeed. But even then, this would not entail that I had not been weighing up relevant issues prior to snatching at one door handle or another. Deliberation is not dependent on any successful resolution of the dilemma over what it is one may be pondering.

9.3 The Accusation of Insincerity

However Peter van Inwagen has another objection to level at the fatalist. He offers an argument to the effect that those who profess not to believe in free will 'either do not mean what they say or they are inconsistent'.

In my view, if someone deliberates about whether to do A or to do B, it follows that his behaviour manifests a belief that it is possible for him to do A—that he can do A, that he has it within his power to do A—and a belief that it is within his power to do B.

He levels an accusation at Baron d'Holbach, whom he takes as a representative example of a disbeliever in free will. For van Inwagen it is 'a proposition as near to being uncontroversial as any philosophically interesting proposition could be' that no one could deliberate about whether to perform or not perform an act that he fails to believe it is possible for him to perform. From this it follows:

... either Holbach never deliberated or else he believed in the case of some pairs of incompatible courses of action that each was within his power. Did he deliberate? Well, of course he did. If he hadn't, he would be as notorious as Pyrrho. A man who didn't deliberate would either move about in random jerks and scuttles, or would withdraw into catatonia ... Therefore he believed in free will, or, at least, in his own free

---

will. Were his beliefs therefore inconsistent? Or should we say that he didn't really believe that there is no free will, but merely said he did? 6

Van Inwagen isn't sure which is actually the case, but either way the fatalist loses: either he is held to be inconsistent or unable to live according to his professed beliefs.

But why should we follow van Inwagen in this? His argument stands or falls on his view of deliberation and how it relates to one's behaviour, specifically the question of how one's beliefs are congruent with one's behaviour. Allowing incongruence would undermine his argument, but he finds this possibility puzzling:

If I am right about deliberation, then there is a very simple reason why one cannot deliberate about what one does not believe ... is possible: when one deliberates, one's behaviour manifests a belief that what one is deliberating about is possible. But couldn't one's behaviour manifest a belief that one doesn't have? This strikes me as a very queer question. 7

I do not see why he should think this so odd. Consider the following example. Despite its somewhat theatrical nature, it bespeaks a very ordinary and understandable fact about human behaviour. You are presented with a challenge in a bar: there is a large sealed glass container inside which there is a rattlesnake. You are told and, we will assume, are completely satisfied that there is no possibility of the glass breaking and the snake escaping, no manner how roughly it is handled or how fiercely the snake might strike at the inside of the container. (You are in fact—we may further stipulate—a scientist whose field of study encompasses precisely this sort of material and its properties.) You will win the offered prize only on the conditions (i) that you place your open palm firmly against the outer glass wall of the container until such time as the snake, who will be encouraged by appropriate agitation on the other side of the container by its owner, lunges at your open hand, and (ii) that you keep it there while it strikes. I do not think we would be very surprised to discover that neither you, nor indeed few others, could win this prize. It is possible that you will keep your hand there, but a much more likely outcome, I think, would be an instinctive reflex pulling away from the glass either before or during the strike. Indeed, there is undoubtedly very good evolutionary advantage accruing to precisely this type of disjunctive behaviour, where one acts completely

7 van Inwagen (1983: 156).
inconsistently with one's firmly grounded and sincerely held theoretical beliefs about the world.

Thus, even if I am wrong about what deliberation really is and, as a fatalist, I am caught exhibiting van Inwagen's revealing and (he would say) self-undermining, deliberative behaviour, I do not believe this demonstrates either inconsistency or insincerity on my part. The ultimate way to demonstrate sincerity would be of course to modify your behaviour on the grounds that your theory of the future had demonstrated facts about your situation that you accept are incompatible with going on as before. Perhaps then the fatalist should just bite the bullet and accept that his doctrine has some difficult consequences. You cannot expect that the truth of your theories will not from time to time occasion some measure of conceptual dissonance. But I suspect that it is very likely that we would not want to accept this. The urge to deliberate in undoubtedly very ingrained, and it is not likely to be shifted by the adoption of a new philosophical stance.

Even if it turned out that deliberation was impossible for a fatalist, we can reflect on what effect this would have on the actual practice of deliberation—or whatever it is that we would now have to call what had hitherto gone by that name. For one thing that is undoubtedly true is that the type of behaviour which we had previously been willing to call 'deliberation' would still be around. It is improbable to imagine that people would stop reflecting on their choices just because of their new awareness of their situation. And what could we say about the earlier behaviour of people? No doubt they thought they were deliberating in the past.

I have been arguing that a belief in a contingently open future is not necessary for deliberative behaviour and against the view that fatalists who do not believe in such a future should stand accused of insincerity in their persistence in such behaviour. However, I noted before that there were two issues: (i) a situation characterised by a belief that real alternatives exist, and (ii) one where such alternatives really obtain. I want to look now at the second and consider whether the fatalist can resist the objection that deliberation is pointless where it is not the case that things may either happen or not happen, as chance or human intervention has it. This will involve us in a discussion of the different attitudes we seem to adopt with respect to deliberation over past choices as opposed to future ones.
9.4 The Pointlessness of Deliberation about the Future

It is often observed that we are all fatalists about the past. This thought is usually offered in the context of a discussion about the plausibility of fatalism about the future. Yet, as Robin Small has noted, the point has some application also to the question whether fatalism is incompatible with deliberation. Small considers the following two claims:

(1) We are all fatalists about the past.
(2) Deliberation about past events is pointless.

He sees the supposed truth of (1) as explaining why (2) is true. As Small acknowledges, this pointlessness is then extended to cover deliberation about the future: 'From this position it is easy to argue that if we were fatalists with respect to the future as well, we would regard deliberation about the events of the future as pointless'. Statements (1) and (2) are being appealed to in order to suggest that the following statement is also true:

(3) Deliberation about past events is pointless.

Recall that Aristotle claimed in the *Nicomachean Ethics* (1139b) that we deliberate only about 'what is future and capable of being otherwise'. 'Capable of being otherwise' in this context means that the events in question are contingent, and with respect to their existence, things may go either way in the sense that they may happen or not happen. Events in the past are an instance where this is not the case: 'Not the gods can shake the Past', as the poet says. Events like this that can go either way must reside in the future (although of course not all future events are like this). We should also note that being an event that may go either way is not sufficient to establish it as an object for deliberation. I do not deliberate about whether I will roll a five with this die (although I may deliberate whether to roll the die at all); nor do I deliberate over whether or not to cause the moon to fly off into space, or make it the case that numbers are real. I may wonder whether numbers really exist, but whether they are or not is not up to me. Such things are rightly regarded as being beyond my control.

---

8 See, for example, Cahn (1967: 96).
10 Aristotle (1941: 1024). The whole passage is quoted above in Chapter 2.
Deliberation assumes the ability to intervene with efficacy. Small observes that while we can value or judge what has happened in the past, we cannot put such judgements into practice.\textsuperscript{11} They are already a determinate part of the past, and any form of deliberation about them seems pointless. But why should this be so? Small attempts to explain:

The usual way of stating our relation to the past is that we cannot undo what has already been done. But the truth is that we can neither do nor undo what has already been done. It is not just that we cannot change the past, but we cannot act upon the past. Given this fact, deliberation about the events of the past is pointless, because no decision to bring about or prevent any past event could be effective. In this respect deliberation about the past is different from deliberation about the future.\textsuperscript{12}

Small’s point is that this crucial difference is grounded in our differential abilities to act on past as opposed to future events. Small is quite right of course about our differential abilities, but only in the absence of backward causation. If backward causation is possible, then we may be able to affect past events in the manner required to ground a legitimate sense of deliberation about them. If this is true, then the fatalist will simply admit that deliberation about the future is no more pointless than deliberation about the past, which, despite what we might ordinarily think about the pointlessness of the latter, may prove to be very much to the point. But even if backward causation turns out to be impossible, the fatalist need not be concerned, though not for the reason that Small advances. Authentic deliberation does not require an actual ability to affect the event about which one is deliberating. All that is required is that the agent \textit{believe} he can affect the event in question. Consider again some of our earlier examples: if I believe (for whatever strange reason) that I can affect the orbit of the moon or the ontological status of numbers, then I can quite licitly deliberate about doing such things. I can’t really make the moon fly off into outer space, but my supposing I can is sufficient ground for my act of deliberation over the matter to count as authentic. In order to have a point, deliberation requires only the appropriate intentional attitude. Being able or unable to enact whatever decision arises out of this process of deliberation is beside the point.

Agents typically believe they can affect future events. Even the fatalist, as we have repeatedly seen, need not deny this. Future event E may in fact only be real

\textsuperscript{11} Small (1988: 29). He notes that the Idle Argument (which I discussed in Chapter 1) works for past events.

\textsuperscript{12} Small (1988: 30).
as a direct result of the agent’s intervention. But it is not just because of this ability to bring about the future event that we can retain a legitimate sense of deliberation for the fatalist. Murderous Tim can deliberate whether or not to kill his grandfather, although he cannot bring about his death if grandfather actually survived. Similarly, Dummett’s chief can deliberate over whether to dance in order to have made the hunters brave. But in these cases, even if it turns out that neither can act on the past, it remains the case that they were deliberating. So the truth of (1) cannot, as Small rightly claims, provide any basis for proving the incompatibility of fatalism and deliberation—but not for the reason Small adduces.

9.5 Concluding Remarks

I conclude then that while fatalism would have far-reaching conceptual consequences for our metaphysical theories, it is unlikely to have profound effects on the ways in which we act in everyday practical situations. This should not in itself lead us to question the validity of our previous arguments. My contention has been that ontological fatalism is a consequence of a particular temporal perspective. This fact is unchanged by any abilities we may demonstrate to accommodate its truth in our ordinary lives.

I alluded earlier to Hermann Broch,13 who observed that fate was stronger than time, in that the final secret of time lay hidden in fate. It would appear, however, that the converse is none the less true: if the ontology of the future is as the global temporal realist perspective of the B-theorist claims it is, then the reality of the future is the ground upon which the truth of fatalism stands.

13 In the first epigraph to Chapter 1.
BIBLIOGRAPHY

We've no free choice, whatever books we read.

Geoffrey Chaucer
Troilus and Criseyde
(trans. Nevill Coghill)

Abelson, Raziel

Albritton, Rogers

Anglin, W.S.

Anscrombe, G.E.M.

Aristotle

Arntzenius, F.

Augustine, Saint

Aune, Bruce

Ayer, A.J.

Baylis, C.A.
(1936) 'Are Some Propositions Neither True Nor False', Philosophy of Science, 3, 156-66.

Ben-Menahem, Yemima

Benford, Gregory

Bennett, Jonathan

Bernstein, Mark

Bertman, Martin A.

Black, Max

Bradley, R. D.

Brams, Steven J.

Broad, C.D.

Broadie, Sarah Waterlow

Brown, Bryson

Brown, Charles D.

Burrell, D.
Cahn, Steven M.  


Capek, Milic (ed.)  

Cargile, James  


Cicero  

Clarke, C. J. S.  

Courtenay, William J.  

Craig, W. L.  

Craig, William Lane  


Davies, Paul  

Denbigh, K.G.  
Deutsch, David

Deutsch, David & Lockwood, Michael

Dickason, Anne

Dummett, M.A.E.


Dwyer, Larry


Earman, J.


(1995a) 'Recent Work on Time Travel', in Steven F. Savitt (Ed.), *Time's Arrows Today: Recent physical and philosophical work on the direction of time*, (Cambridge: Cambridge University Press), 268-310.

Fales, Evan

Faye, Jan

Ferré, Frederick

Fine, Gail

Fischer, John Martin (Ed.)


Fitzgerald, P.


Flew, A.

Fraassen, Bas Van


Frede, Dorothea
Gale, Richard

Gallois, André

Ganssle, Gregory E.

Gaskin, R. M.


Geach, P. T.


Geroch, Robert

Gödel, Kurt

Gorowitz, S.
(1964) 'Leaving the Past Alone', *Philosophical Review*, 73, 360-71.

Grandy, Richard E.

Grant, C.K.
Haack, Susan

Harrison, J.

Hasker, William

Herbert, Nick

Hintikka, Jaakko.

Holdcroft, D.

Horgan, Terence

Horwich, P.


Jacobi, N.

Kenny, Anthony

Kneale, William & Martha

Kvanvig, J.
Lemmon, E.J.

Levi, Isaac

Lewis, David

Lewis, David

Lilley, S.

Linsky, Leonard

Locke, Don


Loizou, Andros

Lowe, Malcolm F.

Lukasiewicz, Jan


Macbeath, Murray

Mackie, J.

Mackie, J. L.  

Makepeace, Peter  

Malament, D.  

Markosian, Ned  

McArthur, Robert P. & Slattery, Michael P.  

McCall, Storrs  


McKim, Vaughan  

McTaggart, J.Mct.E.  

Meiland, J.W.  

Mellor, D.H.  


Morland-Dyke, H.  
(1995)  'Making Sense of Time Travel', *Cogito*, 9, 244-8.

Moyal, Georges J. D.  

Naylor, Margery Bedford  
Nerlich, Graham

Ni, Peimin

Nietzsche, F.

Normore, Calvin G.
(1993) 'Petrus Aureoli and his Contemporaries on Future Contingents and Excluded Middle', *96*, 83-92.

Nozick, Robert

Oaklander, L. Nathan


Oaklander, L. Nathan and Smith, Quentin (eds.)

Oddie, Graham

Olin, Doris

Penrose, Roger

Pike, Nelson

Price, Huw

Prior, A


Prior, A. N.

Putnam, H.


Quine, Willard van Orman

Ray, Christopher

Remnant, Peter

Rescher, Nicholas


Riggs, Peter J.


Rowe, William L.

Ryle, Gilbert

Sainsbury, R. M.
Saunders, John Turk
(1962a) 'Professor Taylor on Fatalism', *Analysis*, 23, 1-2.

Schlesinger, George

Schulman, L.S.

Schulte, Joachim

Seddon, Keith

Segal, Robert

Sharvy, Richard

Sklar, Lawrence
Small, Robin  

Smart, J. J. C.  


Smith, J. W.  
(1985) 'Time Travel and Backward Causation', *Cogito, 3*, 57-67.

Sobel, Jordan Howard  

Sorabji, Richard  

Spellman, Lynne  


Sprigge, T.L.S  

Stein, H.  


Strang, Colin  

Swain, Corliss G.  
Taylor, Richard
(1964a) 'Deliberation and Foreknowledge', American Philosophical Quarterly, 1, 73-80.
Taylor, Richard & Chisholm, R.M.

Thalberg, Irving

Thom, P.
(1975) 'Time-Travel and Non-Fatal Suicide', Philosophical Studies, 27, 211-16.

Thomason, Richmond H.

Thorne, Kip

Trundle, Robert
van Eck, Job

van Inwagen, Peter

van Rijen, Jeroen

Vihvelin, Kahdri

von Wright, G.H.


Vuillemin, Jules

Waterlow, Sarah

Weingard, R.

(1979) 'General Relativity and the Conceivability of Time Travel', Philosophy of Science, 46, 328-32.

Weir, Susan

Wertheimer, Roger

Wheeler, John
White, Michael J.  

Widerker, David  


Williams, C.J.F  


Williams, Donald  

(1954) 'Professor Linsky on Aristotle', *Analysis*, 40, 253-5.

Wilson, H.V.R.  

Wolff, P.  

Wright, Crispin  

Yourgrau, Palle  