

https://dspace.org/

The metaphysics of freedom : time, Kant and compatibilism

Bekker, Duncan

2015-08-20
http://hdl.handle.net/10539/18291

Downloaded from a DSpace repository. https://dspace.org/

MASTERS (MA) RESEARCH ESSAY
DEPARTMENT OF PHILOSOPHY
UNIVERSITY OF THE WITWATERSRAND

The Metaphysics of Freedom

Time, Kant and Compatibilism

By

Duncan Bekker

0708070F

Supervised by Murali Ramachandran

27/05/2015

Abstract: "Does a commitment to the block universe and 'tense-less' B-series time help meet the determinist challenge to free will advanced by the incompatibilist?" Incompatibilism is the position that the free will thesis and the thesis of determinism are not co-tenable. Compatibilism is the denial of incompatibilism. In this essay I examine a theory advanced by Carl Hoefer which aims to establish the compatibility of free will and determinism based upon – what he claims – is the correct understanding of time. I submit that Hoefer's theory faces problems but that his essential insight regarding the relevance of time to the free will debate is important. As such, I analyse various arguments for and against incompatibilism by the lights of a particular metaphysic of time: that given by McTaggart's B-series, Eternalism and the block universe. I suggest that many incompatibilist arguments turn upon an unjustified assumption regarding asymmetries in time and sketch various possible amendments to arguments on both sides of the debate. In particular, I suggest that arguments regarding the challenge to free will from determinism cannot proceed by invoking merely determinism but must deal with the metaphysics of time and causation.

Table of Contents

1 Introduction 1

2 Freedom, Free Will and Determinism 3

2.1 Free Will 4

2.2 Determinism 5

2.3 Definitions 8

2.4 Incompatibilism 9

3 Freedom and the Metaphysics of Time 13

3.1 The two times 13

3.2 The block universe 17

3.3 Freedom from the Inside Out 19

3.4 Trivial consequences 24

3.5 A potential Kantian solution 30

4 Asymmetries, Causation and Costs 37

4.1 Two versions of the Consequence Argument 38

4.1.1 The First Consequence Argument (FCA) 38

4.1.2 The Third Consequence Argument (TCA) 40

4.2 Consequences and assumptions 41

4.3 The ability to render false 46

4.4 The costs of compatibilism 49

4.5 Determinism, causation and the view from nowhere 53

5 Conclusion 56

Bibliography 58

1 Introduction

This essay will be concerned with the apparent challenge to free will from determinism. In particular, it will answer the question whether a commitment to B-series time and the block universe is relevant to meeting the incompatibilist challenge to free will from determinism.

Free will and determinism are often thought to be *prima facie* incompatible. The philosophical literature is replete with arguments attempting to establish this tension. One of the most powerful and persuasive of these arguments has been presented by Peter Van Inwagen whose so-called 'Consequence Argument' seems to provide good justification for the conclusion that no possible deterministic world is a world in which agents have free will. There have been numerous replies to this argument and numerous theories which attempt to establish that determinism and free will are in fact compatible. One of the most intriguing of these replies is a 2002 paper by Carl Hoefer, "Freedom from the Inside Out". Hoefer's theory is heavily indebted to Kant's claim that time – or a distinction between temporal and atemporal 'worlds' – is of vital importance to human freedom. Hoefer does not reply directly to incompatibilist arguments but rather defends an account based upon a particular metaphysics of time. He claims that such an account shows free will and determinism to be compatible in a much deeper sense than has often been thought. Hoefer believes that an account based upon McTaggart's 'tense-less' B-series of time reveals determinism and free will to be robustly compatible.

In light of Hoefer's work, this essay will provide an analysis of the apparent incompatibilist challenge to free will – an analysis performed by the lights of a particular metaphysics of time. I attempt to establish the extent to which concerns about the metaphysics of time are relevant to the free will debate in general and the question of incompatibilism in particular. I submit that the nature of time does indeed play role in the debate. Namely, a B-series metaphysic of time and the related notion of the 'static' block universe reveals a fundamental assumption that may undermine the strength of certain incompatibilist arguments. While Hoefer believes that the revelation of such an assumption allows a compatibilist to mount a positive defence of freedom in deterministic worlds, I believe that this goes too far. An analysis of incompatibilism in the block universe suggests only that incompatibilist arguments are not so powerful as they may intuitively seem. My first substantive conclusion then is that Hoefer is correct about the relevance of time to the problem of free will – a fact seemingly overlooked since Kant first suggested it in his critical

period. However, I also submit that Hoefer's claim to the effect that a metaphysics of time allows a positive *defence* of freedom is not justified. Rather, I claim that both incompatibilist and compatibilist arguments have been slightly off the point – a fact revealed by the connection between time and causation. As such, I suggest various amendments to arguments for both positions and sketch where the philosophical focus should be placed in the future: upon determinism *and* a metaphysics of time and causation.

To this end, the essay will be divided into a number of sections. The first section is this introduction. The second is primarily expository. I present the problem of free will as I understand it and also provide various clarifications and definitions of the terms and positions involved. I then turn to a number of arguments for incompatibilism. The third section deals with Carl Hoefer's theory of compatibilist freedom. I present the distinction between the A-series and B-series and argue that the B-series provides the correct description of the reality of time. I analyse Hoefer's thesis and argue that although his core insight regarding the nature of time is important, the positive account that he provides is problematic and at best incomplete. I sketch a possible amendment to this positive account. In the fourth section I turn to a number of incompatibilist arguments and suggest that, given Hoefer's essential insight regarding the nature of time, they are not so powerful as they appear. In particular, numerous arguments for incompatibilism are based upon a lurking metaphysical assumption about an asymmetry in time – an assumption that is in need of justification. As such, both positive compatibilist arguments of the kind advanced by Hoefer and various incompatibilist arguments are incomplete. This incompleteness turns upon assumptions about time and causation. I submit that the debate has been somewhat off the point. Any potential answer to the question of the compatibility of free will and determinism must take a position on the metaphysics of time and causation. I sketch various such positions.

My conclusions appear in section five and are rather conservative. I submit that concerns about the metaphysics of time are relevant to the free will debate; in particular, it seems that various incompatibilist arguments rest upon unjustified assumptions about time which are in need of further argument. However, I deny that this justifies compatibilism out of hand. I submit that both positions are in need of amendment; that such amendments come at a philosophical cost and that a full accounting of such costs requires a focus upon time and causation as they relate to determinism and its challenge to free will.

2 Freedom, Free Will and Determinism

The debate about free will is so large and diffuse, with so many competing definitions of what freedom is and what kinds of freedom are 'worth wanting' that it is an important for any essay on freedom get really clear at the outset about what the so called 'problem' really is. In *An Essay on Free Will*[1], Peter Van Inwagen, whose wonderfully clear philosophy I will frequently draw upon in this paper, distinguishes between the 'traditional' and 'compatibility' questions. The traditional question is about the fact of our being free – it asks 'are we free?' The compatibility question is about certain theses being co-tenable with our being free – it asks, as just one example, 'if the world is deterministic, *can* we be free?' The questions are of course linked; an answer to the compatibility question, combined with the affirmation of - for instance, determinism - will give us an answer to the traditional question. This being the case, much of the debate around free will has focused upon the compatibility question. It is important to keep the questions separate – a point I will discuss near the end of this section, after the clarifications below.

I suggest that the problem of free will arises given the kinds of answers that the compatibility question has received. That is, the problem of free will is not really about the traditional question. It is *not* that we might or might not be free and we might or might not be upset about this; the problem is not to 'secure freedom'. Rather, the various arguments and answers to the compatibility question seem to suggest that there is something incoherent about the very notion of free will itself. For there seem to be good arguments that free will is incompatible with determinism. Likewise there seem to be good arguments that free will is incompatible with *in*determinism. But presumably worlds are either deterministic or indeterministic, in which case free will cannot exist in any possible world – it is an absurdity. 'So much the worse for free will', one might be tempted to say. But there exist further seemingly good arguments that moral responsibility requires free will and it certainly seems as if the notion of moral responsibility is not an absurdity; after all, we praise and blame and morally evaluate people every day[2]. The *problem* of free will then is that we have good reasons to embrace a notion of moral responsibility, which, we have good reasons to think, depends upon another concept, free will, which we have good reasons to reject. What is going on? A *solution* to this problem is to suggest which class of arguments is flawed and to defend

[1] Van Inwagen, (1983)
[2] Van Inwagen, (2008), pp. 330

one's reasons for claiming they are so flawed. This paper is concerned with arguments of the first class – that is, arguments that purport to establish the incompatibility of free will with determinism.

I will now make some comments about both free will and determinism before presenting a number of terms and definitions that clarify my understanding of various positions in the debate and will be useful throughout this essay.

2.1 Free Will

Free will is a philosophical term of art the definition of which has itself been hugely controversial affair. A review of the literature on free will often suggests that much of the furore turns not upon whether free will is compatible with determinism but what the definition of free will is. Such a state of affairs is regrettable in the sense that it has led some to conclude that the entire debate is nothing more than philosophical squabbling about terms, each camp claiming that only their stipulated account really captures what is contained in the notion of free will.[3] For my purposes, I will distinguish between two broad conceptions that have been popular amongst philosophers[4]. The first is metaphysical and modal and has to do with an agent's abilities, namely her ability 'to do otherwise'. It is the notion that free will has to do with choosing between alternative possibilities: if an agent acts and her act was in some sense necessary – she couldn't have refrained from acting – so that only that very act was possible, then that agent was not free. This captures one platitude about how we naturally think about free will: that if things can only go one way in a situation, then an agent acting in such a situation is not free. But thinking about this fact seems to suggest a second intuition that also finds a natural fit with the notion of free will. It seems natural to say of an agent in such a situation that it was *not up to her* that what happened did happen and so she was not free in the sense that she was somehow disconnected from the facts that obtained; they did not obtain because of *her*. This second conception, focusing not upon counterfactual

[3] See Honderich (2004) for a discussion of the serious problem with this style of argument. The debate about freedom cannot be settled by claiming that one's concept of freedom is the 'correct' one – or that it finds a better fit with what we think is necessary for the attribution of responsibility to an agent. What is needed is a response to arguments claiming that a stipulated definition is compatible or incompatible with other theses regarding such agents and the world.

[4] Pereboom (2001) distinguishes between traditions trading upon what he calls the *leeway* and *source* intuitions regarding free will. Fischer (1994) identifies a similar distinction and discusses it in terms of *regulative* versus *guidance* control.

possibilities, but upon the agent's actual role as the source or origin of her action, has been used[5] to suggest that agents can be free even if they could not do otherwise. I will not deal with arguments for and against this position in much detail during this paper. However, I will occasionally make reference to the notion of an agent's freedom consisting in her originating an action; when I do so I will distinguish this shift in sense.

I will deal with free will on modal and metaphysical grounds. As such, I take it that free will consists in the abilities of agents: the ability to 'do otherwise' or actualize alternative possibilities.[6] More specifically, it consists in the state of affairs in which an agent possesses – with regards to some contemplated act of hers – two abilities: the ability to perform the act and the ability to refrain from performing the act. By **The Free Will Thesis** (henceforth, FWT) I mean the thesis that this state of affairs obtains, with regards to some agents, some of the time.[7] The FWT is hence a thesis about the abilities of agents.

2.2 Determinism

I gloss the essential idea of determinism as the notion that, given the state of the world at some time, the state of the world at later times is fixed as a matter of natural law. That is, determinism is a position that relates states of that world and claims that the relation is one-one, in accordance with one-one laws. A couple of clarifications are required. First, the laws of nature play a crucial role in the concept of determinism. I will remain agnostic regarding what specific account of the laws of nature we should accept. However, it is vital to point out that any notion of determinism requires that there *are* laws. It is the existence of laws that are deterministic that gives rise to the notion of determinism. Therefore, accounts which deny the existence of laws simpliciter[8] amount to a denial of determinism. Positions which provide

[5] Famously by Harry Frankfurt (1969).

[6] Van Inwagen, (1983), pp. 8: 'A person has free will if he is often in positions like these: he must now speak or be silent, and he *can* now speak and *can* now remain silent; he must attempt to rescue a drowning child or else go for help, and he is *able* to attempt to rescue the child and *able* to go for help; he must now resign his chairmanship or else lie to the members; and he has it within his power to resign and he has it within his power to lie.'

[7] For an account of what it is for an agent to have an ability, and for the drawing of a distinction between the active, agential sense of ability and a passive sense of capacity, see van Inwagen (1983) pp. 10. For my purposes, I take the meaning of 'ability' to be given by its natural language usage: an agent 'is able' to perform some act just in case she is able to so arrange her environment (including herself, if it is sensible to speak of 'arranging oneself') so that the conditions for the performance of the act are met and nothing is stopping her performing it.

[8] For instance, Cartwright (1999).

different accounts of the nature of laws – that they are descriptive rather than prescriptive, for instance – will not have any fundamental bearing on a definition of determinism. For, whether the laws merely describe a system of universal regularities or whether they play some deep explanatory or constraining role, as long as the laws hold – that is, as long as they are *laws* – states of the world will be related to each other by these laws.[9] And if the laws are deterministic, the states of the world will be deterministically related. Laws may be expressed by propositions and the conjunction of one or more propositions that expresses a law of nature is itself a law of nature. Finally, it is necessary to note that, under discussions of determinism, we need to carefully limit laws so that they do not trivially express essentially *in*deterministic relations. That is, we must at the very least stipulate that propositions expressing deterministic laws are not disjunctive.

To speak of the 'state of the world at some time' is to seek to capture all facts about the world at some particular instant. For any such set of facts – that is, for any state of the world at some time – there exists a true proposition that expresses these facts. We express the entire set of facts about the world at some time to ensure that there is no lawful interference which would lead to outcomes other than those that the deterministic laws entail. That is, if what was to be deterministically related were not states of the entire world at some time, but rather local states, it would be possible for some other fact, not described in our context but nevertheless governed by deterministic laws to interfere with the particular relation under discussion and so render its outcome other than that predicted by the laws.[10] This would not be a failure of determinism, since the interfering fact would be deterministically related, given a more complete description. The thesis of determinism then must relate complete states of the world in order to ensure that there is no unlooked-for but nevertheless lawful interference. It is also obvious that we cannot allow anything to count as a state of the world that already contains a stipulation of a different state. For example, we cannot allow a proposition P to express a state of the world at time t1 if it also expresses the fact that the world will be in a different state at t2. We cannot allow propositions expressing states of the world to themselves entail different states of the world.

[9] Hoefer, (2010)

[10] See Russell's (1912) famous paper on the failure of a cause being sufficiency for its effect outside of a complete description of context. For another treatment of the issue, see Mackie's (1980) 'INUS' account of causation.

Finally, the arguments I will deal with carefully avoid the identification or even overt relation of determinism and causation – 'determinism is not a thesis about causation.'[11] The intuitively obvious reason for this is that should the world turn out to be indeterministic, this would not imply that there would exist no causation. Or, similarly, how does it follow from the fact that everything is caused that there is only one way things could go? Presumably to answer this question, one would stipulate that it is *deterministic* causation; the point is that determinism is in no way obviously the *same* as causation. Second, and more interestingly, there is the fact that the theories of physics which we identify as deterministic comprise laws that are bi-directional and invariant under time-reversal.[12] Propositions expressing earlier states of the world entail propositions expressing later states, but the converse is also true. However, it seriously strains any acceptable notion of causation to claim that causes and effects are bi-directionally related.

The above is puzzling in light of the tendency in much of the literature to talk of determinism explicitly in terms of causation; as in the phrase 'causal determinism.' In the next section of this essay, I turn to an examination of a paper by Carl Hoefer. He believes that the practice of equating determinism with causation is a result of our natural, common sense view of time. He claims that 'we usually stay in our A-series perspective on the world [and] tacitly conflate determination with causal explanation.'[13] Much of the next section will deal with the nature of time, including explanations of terms such as 'A-series' in the above quote and an analysis of Hoefer's argument. I will return to the topic of the distinction between causation and determinism in greater detail in the fourth section of this paper – after the analysis of the metaphysics of time and Hoefer's argument are complete. And I will argue that such a distinction is of vital importance to the problem of free will. So while it is true that determination and causation are often conjoined in the literature, I submit that such conflation should be treated with care – the relationship is in need of analysis and should not be taken as intuitively obvious. To allay any potential objections regarding my inclusion of such a distinction in my definitions, it is also worthwhile to note that *all* of the incompatibilist

[11] Vehvelin, (2011), "1. Definitions and Distinctions". See also, Van Inwagen (1983); Hoefer, (2004); Earman, (1986)
[12] Hoefer, (2004), pp. 2-3, van Inwagen (1983), pp. 65. Finally: Popper, (1956): 'Classical … mechanics, of continuous media as well as of particles, can describe physical processes only in so far as they are reversible in time.'
[13] Hoefer, (2002), pp. 208

arguments with which I shall deal carefully treat determinism and causation as separate issues.[14]

It is useful briefly to clarify the relationship between free will and determinism as both have been discussed so far. If the free will thesis is a thesis about the abilities of agents with respect to their contemplated acts, and determinism is a thesis about propositions expressing states of the world and the laws of nature, how exactly are they to be understood as related? If an agent S has the ability to, for instance, raise her hand at t, then she has the ability to render the proposition, S raised her hand at t, true. This proposition can express a fact, part of the 'world state' proposition which shall be one of the propositions that the thesis of determinism shall be about. Agents have abilities relating to propositions just in case their abilities have consequences for the facts expressed by propositions, which, of course, they do.

2.3 Definitions

The Free Will Thesis (FWT) is the thesis that agents, on some occasions, have two abilities with regards to a contemplated act of theirs: the ability to perform the act and the ability to refrain from performing the act. **Determinism** (DT) is the thesis that if a proposition P expresses the state of the world at some time, the conjunction of P with a proposition which expresses the laws of nature (L) entails Q, a proposition that expresses the state of the world at some other time. **Indeterminism** is the denial of determinism; assuming the same terms as above: (P ^ L) does not entail Q – of course, on indeterminism, neither does (P ^ L) entail not-Q. **Incompatibilism** is the thesis that FWT and DT are incompatible; one cannot coherently affirm both theses. **Compatibilism** is the denial of incompatibilism. **Libertarianism** is the affirmation of both incompatibilism and FWT. Libertarianism therefore is the denial of DT and the affirmation of indeterminism. **Hard determinism** is the affirmation of both incompatibilism and DT. Hard determinism therefore is the denial of FWT. **Soft Determinism** is the affirmation of DT and FWT. Soft determinism therefore implies compatibilism.

[14] One important example of this explicit separation is Earman's (1986), pp. 5 statement that to invoke causation in a discussion of determinism is to 'explain a vague concept – determinism – in terms of a truly obscure one – causation.' Another is van Inwagen's (1983), pp. 65 definition of determinism in which he notes that 'the horrible little word "cause" does not appear in this definition. Causation is a morass into which I for one refuse to step foot.'

This essay will deal with incompatibilism and a potential soft determinist defence. It is vital to note that while libertarians and hard determinists are incompatibilists, the arguments they advance are not arguments *about* the incompatibility of freedom and determinism. They consider that question as having been resolved and they are arguing about whether or not we have free will. When I refer to incompatibilists, I mean those philosophers who are attempting to settle the thesis of incompatibilism - not those who are arguing for an incompatibilist thesis.

2.4 Incompatibilism

When one examines the relationship between the free will thesis and the thesis of determinism one soon recognizes what appears to be a *prima facie* conflict between the two. For the free will thesis, in attributing two abilities to agents, seems to suggest that there are at least two ways that things could go, given an agent's choice as to which ability to exercise. If an agent exercises the ability to jump up and down on the spot, then there seems to be a way the world could be that includes this happening. But, if she is to be free, then she could also refrain from such behaviour and then there would be a way the world could be that does not include this fact. There are two ways things could go, dependent upon which ability a free agent exercises. But determinism immediately seems problematic for such a view. For determinism, at its most fundamental, captures the idea that there is only *one* way that things could go, given the world's being governed by laws.

This intuitive view of the tension between free will and determinism can be used to construct an argument based around a natural representation of the world that we may call, alluding to Jorge Luis Borges' famous story, 'The Garden of Forking Paths.'[15] Such an argument might run as follows:

> CI1) An agent is free only if she is able to actualize genuinely open alternative possibilities.
>
> CI2) For an agent to actualize genuinely open alternative possibilities, there have to *exist* such possibilities.

[15] Van Inwagen, (2008), Chapter 12. See also, Howwich (1987) pp. 25-28, for the 'tree model of reality'.

CI3) If determinism is true, then there is only one open possibility, the one that is a lawful continuation of the world's history.

CI4) So, if determinism is true there do not exist *alternative* possibilities – the only possibility is the actual one.

CI5) Therefore, if determinism is true, agents cannot be free.

Let us call this argument an argument for classical incompatibilism and note that it is founded upon a particular notion of reality, one that relates to the 'open-ness' of possibilities in the future. While the argument has much intuitive strength it is not quite so complete as it may first seem. First, it is grounded upon a particular model of reality – a model that may well turn out to be false (I will say much more about this in the section to come). It also trades heavily upon the fact that things can only go one way in accordance with laws. Many philosophers, whom I will call classical compatibilists, have worried about this – worried that just this fact is not enough to ground incompatibilism. For it does not seem exactly clear how the argument is supposed to exclude counterfactual possibilities. These possibilities do not obtain in the sense that seems to be required by the model of a garden of forking paths, but how does the argument show that they *could not*? The worry is that while it is true that things happen one way in accordance with laws, *they might have gone another way*. After all, why couldn't they have gone another way, if all that is required is lawful continuation? And this might buy room for agents to be able to do otherwise.

Such compatibilists, who offer a counterfactual reading of the abilities of agents, claim that it is correct to say that, *if* an agent S had chosen to do otherwise, then she would have done otherwise. She didn't, but she could have; and this is compatible with determinism. That is, if something had been different - the agent's choice - then a lawful continuation of the world would have included the different facts dependent upon that choice. Such a reading both seems natural and deeply strange. It is natural in so much as it seems to offer a grounded and common sense view of what would have happened if we had chosen differently. It seems correct to grant determinism, that things only go one way in accordance with law, but point out that there does not seem to be anything unlawful about my having made a different choice – how is that *closed*, in the garden of forking paths? Making a different choice does not seem unlawful in the sense that, for instance, travelling faster than light would be. That things go one way does not mean that they cannot have gone another, on such a reading. The classical incompatibilist argument describes a deterministic world

without alternative possibilities but it doesn't seem to show *why* agents couldn't have counterfactual abilities. The strangeness of such a compatibilist reply surfaces, however, when we examine the exact process by which we could have chosen differently. For this process too, is *already* determined, as it were. There has been much back and forth around this point and the correctness of a counterfactual reading of the agent's abilities. At least part of what has emerged is another argument for incompatibilism – an argument that seems to offer an incredibly strong challenge to free will if determinism is true. The argument places its entire focus upon the past and the laws and the way agents relate to these facts. And it attempts to deal with the worries that may undermine the counterfactual reading.

If the Garden of Forking Paths relies upon a description of a deterministic world without freedom then the Consequence Argument is the *explanation* of that world. It aims to explain why agents cannot be free in such a world despite the compatibilist's counterfactual analysis. The Consequence Argument aims to establish a kind of *necessity* relating to facts about the past and facts about laws and facts about how the laws and the past relate to our present actions. Such necessity – although it is not strict metaphysical necessity – purports to show why any compatibilist understanding of agents' abilities to do otherwise must be mistaken. For, in meeting the Consequence Argument they must claim that agents have powers that they necessarily, in some sense, do not have. Van Inwagen presents the core idea of the Consequence Argument thus:

> If determinism is true, then our acts are the consequence of laws of nature and events in the remote past. But it's not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.[16]

This argument emphasizes the apparent necessity of the past and the laws of nature. The necessity is not metaphysical but grounded in a necessity relative to *us* – it is not 'up to us' what went on in the remote past and it is not 'up to us' what laws govern the world. But these things over which we have no control entail that we act precisely as we do in the present moment and so it is not up to us that we so act. If it *were* up to us whether we so acted, if we could do otherwise as the compatibilists claim, then the past and the laws of nature would

[16] Van Inwagen, (1983), pp. 56; See Lamb, (1977) for a very similar argument.

have to be up to us too. But clearly they are not; no human being has control, power or the ability to affect the past and the laws of nature.

The Consequence Argument is often taken to give what might be called a master argument for incompatibilism. It has certainly convinced many, perhaps the majority, of the philosophers who have come across it while considering the incompatibility of free will and determinism.[17] Should it convince us? I submit that, while the argument is powerful and is neither obviously invalid nor unsound, we should be wary of finding for incompatibilism merely on the strength of this argument. It conceals a vital metaphysical assumption that is in need of support. It is to this argument that I will turn in the fourth section of this essay. I delay my analysis and discussion of the Consequence Argument until then because I will make use of a number of concepts and arguments that are themselves in need of analysis and commentary. Particularly, a tense-less conception of time, and an argument to the effect that a commitment to such a conception allows one to formulate a new kind of compatibilist thesis. So, with the understanding that I will return to the Consequence Argument and the thesis of incompatibilism, I turn to a consideration of time and its relevance to the free will debate.

[17] Van Inwagen, (2004), pp. 350

3 Freedom and the Metaphysics of Time

In his 2002 paper, Carl Hoefer presents a rather intriguing thesis arguing that the apparent threat to free will from determinism is illusory. He claims that what endangers our free will is not the physical theory of determinism but rather the 'unholy marriage' of that theory with a mistaken view of time; a view of time that is in fact in direct tension with physics. His basic claim is that 'given the proper understanding of time, we will see that freedom and determinism are compatible – compatible in a much more robust sense than has ever been thought possible.'[18]

3.1 The two times

What is the correct understanding of time? When we ordinarily think of time, one of its most striking characteristics is the seeming difference between past and future. Not simply that the past and future seem 'separate' or go in different directions but rather that they are robustly dissimilar. There are various accounts of why this is so; and philosophical debates about the 'arrow of time' are alive and well.[19] I submit that the fundamental reason for our so conceiving of time is phenomenological: there is an irreducible feeling of 'moving into the future.'[20] The future feels open, while the past, the place we are 'leaving' feels closed, finished, done. We naturally think of our passage into the future as conditioning it, making it be and be how it will be and thereby fixing it. Our natural understanding of time is therefore two-fold – involving a distinction between past events and future events and a sense of temporal motion, a flowing *now*, which explains the distinction. There is a difference between the past and the future; past events are fixed in some sense and so beyond the scope of our ability to influence events. The future, on the other hand, is not fixed, is open in some sense that allows events in the future to be affected by us. Further, we think that there is a *now*, a flowing temporal instant that – as it 'moves' through time – is itself responsible for dividing the future and the past in this way. The *now* makes future, open, events *become* and so fixes them and renders them beyond our reach. It seems to us that the ontology of time and of events turns upon a temporal flow - a coming into being conditioned by time itself moving

[18] Hoefer, (2002), pp. 202
[19] Horwich, (1987), pp. 15-18
[20] Horwich, (1987), pp. 33-35

into a future that is not yet real. And *a fortiori*, it seems to us that events in the past are very different, ontologically, from events in the future.

However, things are not as they seem. John McTaggart noticed that there are two ways in which we order events in time, and used these different means of ordering to construct two series of time, the A and the B. He says: "Positions in time, as time appears to us *prima facie*, are distinguished in two ways. Each position is earlier than some, and later than some, of the other positions. And each position is either past, present or future. The distinctions of the first class are permanent, while those of the latter are not."[21]

Here is how I understand the distinction. We may order an event relative to *now* and say of it that it is in the past, or in the future, or is present. Or we may order an event relative to another event and say of it that it is earlier than, later than or simultaneous with the other. If we systematically order events relative to *now*, we construct a tensed (that is, impermanent) series. He called this the A-series. If we systematically order events relative to other events, we construct a static series. He called this the B-series. The important difference between the permanence of the B-series and the impermanence of the A-series can be drawn out as follows. The events we order on the A-series do not keep their temporal location or position while those we order on the B-series do. If an event is ordered on the A-series as past, very soon it will be ordered as *more* past. That is, if an event is ordered as five minutes in the past on the A-series, very soon (or something like 'soon') after this specific ordering the fact by which we ordered it does not obtain. The A-series is changing. On the other hand, suppose we order an event on the B-series as earlier than another event. Suppose we order it as five minutes earlier than another event. This specific relation shall always obtain between these two events on the B-series; they are forever ordered in that way. The B-series is static.

McTaggart used the distinction between the A-series and the B-series to argue for the unreality of time. Very briefly, his argument proceeds as follows. Time requires genuine change. Only the A-series can account for genuine change, the B-series cannot. So time requires that events be ordered on the A-series. If events are located on the A-series, then every event has the properties of being past, present and future. There is a contradiction in predicating all of these properties of any event. Ordering events on the A-series results in a contradiction. So, the only account of time that accurately captures change, a requirement for any account of time, results in a contradiction. Therefore, time is not real.

[21] McTaggart, (1908), pp. 458;

Granting the plausible assumption that time requires change, the argument can be divided into two theses. The B-series cannot account for change and the A-series results in a contradiction. Since Hoefer will take the B-series to give the *correct* view of time, we need to answer the challenge of the first thesis. That is, show that there is something wrong with the notion that the B-series cannot account for change but also affirm the contradiction that the A-series supposedly gives rise to. Again, very briefly, I present an example of the sort used by McTaggart in establishing the B-series' inability to account for change. Imagine a hot fire poker which proceeds to cool in the period between t1 and t2. It is hot at t1 and cool at t2. McTaggart thinks that this does not make any real change because it will be true throughout the history of the world that this poker is hot at t1 and cool at t2. Those facts are eternal and so McTaggart argues, cannot account for a real change in the world.

For brevity's sake I have presented the argument in a way which makes it reasonably clear where his mistake lies. It is in assuming that genuine change requires that the sum total of facts at some time be different from the sum total at another. When we look at this requirement carefully it becomes obvious that this is to assume that genuine change *means* A-series change; McTaggart begs the question.[22] It is true that the facts about the poker at t1 and t2 never change, but that in no way implies that the poker does not change. The poker changes precisely insofar as it is hot at t1 and cool at t2 and *that fact* – that the poker changed – is also timeless, which is just as it should be. Surely we want to say 'It is a fact that the poker changed. And a timeless fact at that, there was no time at which it would be correct to say that the poker did not change.' Only the B-series allows us to do this. McTaggart mistakenly identifies the subjects of change as events and not objects and so mistakenly concludes that the B-series does not characterize change.

On the other hand, I think McTaggart is correct in claiming that locating events on the A-series results in a contradiction. Here is a radically simplified way of showing this. Events are ordered on the A-series by virtue of their being past, present or future. These properties are incompatible. But, by virtue of the change with which the A-series itself presents (the future becoming present, the present past, the past becoming more past and so on), every event instantiates every property in the A-series. The A-series therefore results in a contradiction.

[22] Dyke, (2002), pp. 139

There is a potential objection here that must be dealt with immediately. A friend to the A-series may be likely to claim the following. There is only a contradiction if an event in the A-series simultaneously possesses these properties, but this is not necessary for the existence of the A-series. Perhaps the properties are had in succession such that for some event E, we may say that E is future at t1 and E is present at t2 and E is past at t3 and these are perfectly compatible properties. This seems promising until one realizes that such a move in fact eliminates the A-series. Ordering E as future relative to t1 is in fact ordering E on the B-series. E's being future at t1 is a static property, the phrase 'being future' on such an account actually comes out as 'is later than'. In making this objection, one abandons the A-series to avoid the contradiction – which is, of course, precisely what the argument against the A-series says we must do.

There is another important point to be drawn from the failure of such an objection. The properties of being past, or future, or present possessed by A-series events are deep. They make an ontological difference. The failure of relativization helps draw this out. What makes a past event past, on the A-series is that it has the property of *being past*. Its pastness matters; this is why there is a contradiction in attributing all these tensed properties to a single event. When we attempt to solve the contradiction by making the properties relative to some fixed point, we get rid of the very property by virtue of which it was ordered on the A-series – for on the A-series, events are fixed relative to a changing point, the *now*. To be committed to the A-series is to be committed to the property *being past* and to be committed to the notion that such a property grounds an ontological difference between events with this property and events which lack it.

It is worth noting that it is not the case that giving up on the A-series means giving up on A-series talk. After all, it may be objected that we use A-series terms such as 'now' and 'is past' every day and no contradictions or confusions seem to be implied. Granted, but a commitment to the B-series as giving the correct metaphysical description of time does not imply *elmininativism* about the A-series or that we should eliminate our usage of tensed language. After all, McTaggart suggests that we do in fact order events in two ways; his argument is just to the effect that only one of these orders can characterize the real nature of time. Realizing that, we merely need an account of how A-series terms function. We must just be aware that such tensed language does not denote some underlying reality – some ontological thing or property *now* or *past*. Given the B-series, these phrases come out as indexical. But of course indexical terms are perfectly proper terms. The sentence 'It is *now*

the case that P' glosses the notion that P obtains at the B-series location of the sentence's utterance. The tensed locution 'now' is exactly analogous in function to the indexical first person pronoun in 'I am hungry' – the 'I' functions to denote that it is the speaker of the sentence who is hungry. It is not the case that the B-series implies that there is something wrong with everyday language, only that we should analyse such language carefully.

3.2 The block universe

The B-series has been used (since Minkowski first did) to provide models of our world which represent it as a four dimensional 'block'. The commitment to such a picture of reality is variously called four-dimensionalism, Eternalism or the block universe. On the block universe view of the world, reality is described in a manifold consisting of three spatial dimensions (x,y,z) and a temporal dimension (w) given by the B-series ordering of the events of that world[23]. The events themselves are unchanging on the B-series but as we trace the properties had by objects across their 'world-line', the objects change. The world-line of an object is given by the complete set of four-dimensional co-ordinates possessed by the object throughout its history. To speak roughly, time becomes a geometric property on the block universe.[24] Just as we would not say that there is any ontological difference between things located at different (x,y,z) co-ordinates in the universe – they all exist, they are just exist at different co-ordinates – we say the same thing about things located at different (w) co-ordinates. When we speak of a 'time slice', we mean a four-dimensional cross section of a region of the block. Such a time slice provides a description of all the facts located at a particular value of (w).

It is worth briefly noting here that the picture becomes significantly more complicated when we take the theory of relativity into account. Given that the speed of light limits the possible speed at which anything can propagate, a caveat arises with regards to our definition of determinism. I mention this just to pre-empt any objections that determinism has a fundamentally different character in the block universe; a requirement for analysing the arguments to come because Hoefer makes occasional use of relativistic notions in his theory of freedom. Given the speed of light it must turn out that there are regions of the block universe 'space-like' (rather than time-like) separated from other regions. Time-like

[23] Savitt, (2013), "Presentism, Possibilism, Eternalism"
[24] Sider, (2001), Chapters 1-4

separation couches the idea that tracing the (w) co-ordinates of light in a region (as it propagates outwards) will eventually link it to another region. Regions space-like separated are disconnected simpliciter along this dimension. The problem, of course, given our definition of determinism covering 'states of the universe at times' is that a proposition expressing a state of the universe which is space-like separated from another does *not* entail the other in accordance with laws – simply because there can be no physical interaction between them and so nothing for laws to cover. The solution is to merely take the relevant propositions for determinism to express states that are never space-like separated; and to translate propositions expressing states of the universe into propositions expressing states of the world relative to frames of reference.[25] The thesis of determinism functions exactly the same way under such translations and for the most part I will continue to talk of states of the world and time-slices given the understanding that they *can* be translated into propositions appropriate for relativity. Relativistic considerations also provide another reason for thinking physics incompatible with the A-series, if more reason be required.[26]

It is finally worth noting the distinctly Parmenidean character of the block universe. Parmenides famously claimed that the world was an unchanging whole.[27] Many of the contemporary objections to this claim were analogous to McTaggart's problems with B-series time in that it cannot account for change. Once it has been established that it is not events, but objects, that are the correct locus for attributions of change it seems plausible to suggest that such worries about change need not undermine the thesis. But the 'unchanging' world still remains an alien one. On the block universe, if it could be viewed as it really is[28], a four-dimensional object, it would truly be a static whole in which, to invoke the language of the pre-Socratic philosophers, nothing *becomes* but everything *is*.[29] I will say more about the consequences of this, particularly as they relate to our notions of causation, in later sections.

[25] For various rather technical discussions see Earman, (1986), Chapter 4

[26] The reason is the following: relative to various frames of reference (given by the inertial motion of an observer and the limitation of the speed of light), different events will come out as simultaneous – that is, simultaneity is relative: there is no absolute now. If we are to be Presentists and accept the A-series and the 'now' as grounding existence, we must then accept ontological relativism – a seemingly absurd idea. See Savitt's (2013) treatment of the 'Rietdijk-Putnam-Penrose' argument.

[27] Gallop, (1984)

[28] That is, from the perspective which Nagel (1989) called 'the view from nowhere.'

[29] Mellor, (1985), pp. 128

3.3 Freedom from the Inside Out

Hoefer makes use of the above notions - B-series time and the block universe - to present his case for freedom. The metaphysics of time given by the block universe is the 'proper understanding of time' that he claims allows us to conclude that free will and determinism are robustly compatible. The insight that does serious philosophical work for Hoefer is the B-series' commitment to events, no matter where they are temporally located on the B-series timeline or in the block universe, having the same ontological character. They are all equally real and nothing is entailed about the properties of an event simply by virtue of it being past. The 'events of 1000 years in earth's future are, in terms of reality or existence, no different from the events (*now*) of your reading these words, or the events of last week.'[30] The A-series intuition that the past is fixed or done in a way that is different from the future is just not true. When this is combined with the fact that determinism relates the states of affairs comprised of these events there is 'no reason to think of past → future determination as more important or real than future → past determination. And, even more to the point, one can equally view a set of events in the *middle* as determiners of both past and future events.'[31]

The point is well made as it relates to the thesis of determinism. The bi-directionality of determinism is granted by van Inwagen and other champions of incompatibilism. But, I think, Hoefer is claiming that this feature of determination has not been given its due. Determinism tells you that states of the world at times determine all the rest. But determinism fundamentally does not tell you *which* state, or *which* time slice, determines the rest. It remains entirely agnostic as to where relations of determination 'begin'. It is here that Hoefer's claim that the challenge to free will has not come from physics and determinism but from an 'unholy marriage' of determination with a further metaphysical picture about time seems to hold water. For, given determinism, one is entirely justified in viewing absolutely any slice as determining another. But incompatibilists who trade upon the thesis of determinism have not realized this – for them, the past pushes us around. Van Inwagen and others have recognized the bi-directionality of determinism but they have implicitly assumed what is then a vital feature of their argument: the status of past time slices. That is, while incompatibilists have granted bi-directional determinism, they have viewed this as a mere curiosity grounded upon the fact that the laws of physics are time-reversal invariant – they do

[30] Hoefer, (2002), pp. 205
[31] Hoefer, (2002), pp. 205

not care which way they are 'run'[32]. In deference to this fact, they have defined determinism in a way which fits with physics, but they have not thought that this has any serious consequences for their arguments. For they have assumed that, while the laws are bi-directional, there is an important asymmetry in what the laws cover: events. They have assumed that there is a real asymmetry in the properties of past and future events. But the B-series and the block universe reveals this to be false – the temporal location of events has no bearing upon their ontological properties.

As Hoefer notes, what we worry about is that a *past* time slice determines the present one, the one in which a contemplated act will take place.[33] This is often made explicit by incompatibilists who claim that, given determinism, the facts of the past entail all the truths about the future. But determinism on the B-series shows that it is equally true that the present time slice determines both future and past time slices. That is, the facts of the future equally entail all the truths about the past. And it is upon this relation that we should focus – for, given the above, facts about the *present* entail all the truths about the future and the past. That is the essence of Hoefer's thesis of freedom from the inside out. **Freedom from the inside out** is the thesis that:

> We are perfectly justified in viewing our own actions not as determined by the past, nor as determined by the future, but rather as simply determined (to the extent that this word sensibly applies) by ourselves, by our own wills. In other words, they need not be viewed as caused or explained by the physical states of other, vast regions of the block universe. Instead, we can view our own actions, qua physical events, as primary explainers, determining—in a very partial way—physical events outside ourselves to the past and future of our actions, in the block. We adopt the perspective that the determination or explanation that matters is from the inside (of the block universe, where we live) outward, rather than from the outside (e.g. the state of things on a time slice 1 billion years ago) in.[34]

[32] Popper, (1956): 'Classical … mechanics, of continuous media as well as of particles, can describe physical processes only in so far as they are reversible in time.'
[33] Hoefer, (2002), pp. 206
[34] Hoefer, (2002), pp. 207

There are two important features of the above account that need clarification. The first is the notion of our being 'justified' in viewing our actions as free, not explained by vast states of the block universe. The second is the notion of our actions determining future and past time slices – and, particularly, determining them in only a 'very partial way'.

Hoefer says almost nothing about the first notion – relating to justification – in his paper. I submit that it is important to get clear about what it is that Hoefer is claiming we are justified in doing. We are *not* justified, merely from Hoefer's account, in concluding that we are free. Such a defence of freedom requires an account of how it is that agents can be or come to be free at all; that is, an account of how agents can come to have the abilities that the free will thesis attributes to them. This is a positive account of freedom and needs to be argued for from the nature of agents. Hoefer is not providing such an account or defence. But it is also vital to note that such an account is not what is really at stake in arguments which purport to establish incompatibilism. Incompatibilism is not the thesis that we are not free – hard determinism is; and although hard determinists are incompatibilists they argue in essentially different ways. Hard determinists assume that the 'compatibility question' referred to in my introduction to the problem of free will has been clearly answered. They then argue that determinism is *true* and hence that we do not have free will. Incompatibilists arguments are not about whether or not we have free will, they are about the apparent absurdity that arises on the assumption that we *are free* and that we *are determined.* The basic cartoon argument is that freedom requires us to have alternative possibilities related to the abilities we have when we contemplate an act at a time. But the past and the laws have strictly determined the state of the world at the time in which we act so that there are no alternative possibilities. The above is the kind of classical incompatibilist position that invokes the idea of a garden of forking paths. Hoefer's point about justification relates to *that* argument and it is well made.

Given the B-series, nothing about past events can ground the notion that it is only the past that determines the future – nothing suggests that past events have special properties. And given the definition of determinism, nothing about the relations between events can ground the notion that the past determines the future. But then what *possible* other reason could we have for thinking ourselves unfree? We can put the point this way: imagine only the time slice at which I act (if you like, imagine a possible world in which only this time slice exists) – is there any reason, given just this time slice, that my act is not free? If the answer is 'yes', then the incompatibilist begs the question, for, in such a case the reason for worrying

that we are not free would not be determinism (perhaps it would relate to indeterminism).[35] Her move is clearly to reply that the above misrepresents the picture, because, in the real world, there are other time slices and determinism shows that there are concerns about *prior* time slices that have effects for my freedom. But what Hoefer has shown is that such concerns are mistaken – there is nothing about the past that renders it 'fixed' or 'done' or 'closed' simply by virtue of being past and nothing about determinism that suggests a unique temporal direction. Neither determinism nor the past gives us reasons for preferring any time slice rather than any other; so in a very real sense, we are justified in viewing the time slice in which I act as primary and so are justified in taking ourselves to be free at that time slice. That is the view we are 'justified' in taking up on Hoefer's account. Concerns about the pastness of states of the world do not preclude free will. Pastness and the laws do not do any work for the incompatibilist. But does Hoefer's thesis not claim that agents can do work relating to the laws and the past?

The obvious incompatibilist response to Hoefer's thesis up to this point is to invoke the powerful intuition which underlies the Consequence Argument: being free at this time slice entails having the ability to influence the past. Hoefer notes that this is the really serious concern with which his account must deal[36] and his solution brings us to the other important claim in his setting out of the thesis of freedom from the inside out: the 'partiality' of determinations. I suppose we may put the point as follows: Hoefer's essential insight has to do two things. First, it has to show that there is something wrong with thinking about the past and the laws themselves as giving only past → future relations of determination; this is what I claim, in the above, that he does rather well. Second, however, it also has to show that there is something wrong with taking future → past relations of determination to be suspect – which suspicions may be glossed as founded upon notions of causation.

Hoefer's project can therefore be broken down into two theses. In the first part he notes the ontological equality of events on the B-series and argues, correctly, that this has serious consequences for the supposed incompatibility of freedom and determinism. Philosophers have been arguing under the assumption of a time asymmetry that does not exist. The second part of his thesis consists in a positive defence of the claim that free will in

[35] Of course, we may worry about the lack of any sort of intelligible process by which an agent could deliberate or originate her action in this single time slice world. But the point is that it is not worries about *determinism* that ground this objection. The incompatibilist about determinism needs to focus on the relation between determinism and freedom, not the relation between freedom and indeterminism.

[36] Hoefer, (2002), pp. 207

the block universe does not involve backwards causation. In presenting this defence Hoefer introduces increasingly technical and subtle distinctions and eventually, I argue, demotivates his thesis as a whole by seriously undermining the simple and powerful insight that grounds the first part of his theory. I offer the following brief sketch of Hoefer's argumentative strategies and some comments which I will deal with in more detail in what is to follow. Hoefer draws a distinction between causation and constraint – constraints coming out as roughly the kind of logical entailment relations discussed by determinism. Hoefer claims that our choices and actions place constraints upon what the past and future are like but do not thereby cause, or causally explain, the past. Such a move indicates that part of Hoefer's strategy is to deny serious causation. However, given his analysis of constraint, it turns out that while our actions shall have very little consequence towards the past, the same is also true of the future. And this seems problematic. To account for this problem, Hoefer then makes what I think is a bad move – one which undermines his entire thesis. He reintroduces causation, a temporally asymmetric relation, and bases it upon physical phenomena that are not explicable in terms of the A-series. That is, he bases his notion of causation upon various '*real'* (non-subjective) physical asymmetries: the thermodynamic asymmetry, the radiative asymmetry and the fork asymmetry. In so doing he undermines much of what I think is good in his paper. For his basic idea - the idea that I like - is that the incompatibilist challenge to free will has come not from determinism but from the combination of determinism with mistaken assumptions about asymmetries in time. However, in re-introducing causation grounded upon what are *real* asymmetries he seems to both undermine his original argument *and* to present a perfect objection to the incompatibilists.

Jason Brennan, in a response to Hoefer, argues that either "Freedom from the Inside Out" is incomplete or it is irrelevant. [37] It is incomplete insofar as Hoefer does not present an argument to the effect that the causal asymmetry he re-introduces is not a threat to free will. I think that this is correct and will briefly sketch the kind of response that I think Hoefer should make – a response Kantian in character given that so much of Hoefer's paper already presents with a distinctly Kantian bent in the separation of the atemporal block universe from the tensed A-series in which we live our lives. Brennan argues that Hoefer's paper is irrelevant in so far as it turns out that causation, not determinism, is the real threat to our free will – and so Hoefer's account, in focusing upon determinism, is irrelevant to the free will debate. I am suspicious of this point. My contention is that Hoefer's core insight regarding the symmetry

[37] Brennan, (2007), pp. 213-214

of events on the B-series highlights an essential gap in the consequence argument. While Hoefer thinks that this gap allows a serious defence of compatibilist freedom I will argue for the more moderate claim that the debate has been mis-focused and that incompatibilist arguments are themselves incomplete. That is, I submit that analysing incompatibilist arguments with a B-series metaphysic of time both shows what is wrong with them and also shows what should be done about strengthening them. This will be my project in the fourth section of this essay. So I both agree and disagree with Brennan regarding his claims of 'irrelevance'. To the extent that Hoefer's insight is supposed to ground a positive account of freedom, he does not make much headway. But to the extent that it allows us to clarify the debate between compatibilists and incompatibilsts, it is important. Let us now turn to the second part of Hoefer's thesis, that freedom from the inside out does not commit one to backwards causation.

3.4 Trivial consequences

Hoefer begins by claiming that his inside-out account does not entail that our actions have 'causal consequences towards the past … [but rather] our freely chosen actions place *constraints* on what the past and the future can be like, but the constraints are astonishingly weak, both towards the future and (especially) towards the past.'[38] I am immediately wary of three potential problems. First the move from causal talk to talk of constraints; in claiming this he is admittedly stressing that determinism and causation are separate notions and appealing to the fact that determinism proceeds in terms of entailment – but we must ensure that his account of constraints fits with various platitudes about our actions; which brings us to the second concern. The second problem relates to the posited 'weakness' of the constraints. Depending on how weak such constraints are on his theory, it might turn out to be incapable of capturing an obvious point about our actions: that they make a difference. If the constraints are so weak as to have almost no effects then in what way has Hoefer defended the notion of free actions as we conceive of them? If the constraints are too weak, Hoefer will have defended a thesis on which 'our freedom to act is the freedom to do very little.'[39] Finally, his parenthetical qualification which claims that our actions are '*especially*' weak towards the past seems strange given that his theory has, up until now, been entirely

[38] Hoefer, (2002), pp. 209-210
[39] Brennan, (2007), pp. 207

focused upon symmetries: the symmetry of the ontological status of events in a block universe and the symmetry and time-reverse invariance of deterministic laws. In fact the only asymmetry noted by Hoefer is that found in the temporally embedded A-series perspectives of human agents. The introduction of what appears to be a commitment to serious physical asymmetry half way through his account requires scrutiny.

What are constraints? Hoefer asks us to assume that a human action is 'a physical event *type* that has innumerable instantiations at the microphysical level.'[40] The idea is that the kind of absolute reduction entailed by theories like the mind-brain identity thesis fail and we are instead left with the weaker notion of supervenience. On which, I think Hoefer claims, the physical portion of a human action has a wide variety of microphysical supervenience bases which count as 'good enough' for the macrophysical action to have obtained. Supervenience is a notoriously vague subject but I think that we can grant Hoefer's assumption as long as we are not under the sway of a very strong theory of event identity. A paradigmatic gloss of supervenience is the claim that there can be no change in the supervening properties without a change in the supervenience bases of those properties. So, if the macrophysical supervenes upon the microphysical then there can be no macrophysical change without a microphysical change. All that Hoefer is claiming is that there are many microphysical ways to, for instance, raise my arm. Unless we are insisting upon a theory by which we identify and differentiate events only upon their microphysical properties, then it certainly seems true that, for example, switching the position of two microphysical particles in my arm doesn't really make a difference to calling the event 'the event of my arm's being raised'. Assuming this is the case, Hoefer then claims that my act of raising my arm entails only that some one of these innumerable microphysical states obtains.

His point here is a good one – at least as it relates to the moderation we should use with regards to ascriptions of ability. My act and its context are only a tiny part of the whole time slice in which the act occurs; that is, it is not the case that the characteristics of the entire time slice follow from my action, the act only determines a very trivial region of that time slice. And the same is therefore true of the time slices immediately to the past and future. As Bertrand Russell first argued (and I mention in my introduction to determinism), no event by itself ever guarantees efficacy with regards to the bringing about of another. Striking a match does not guarantee the match's ignition until we have so enlarged our description of the

[40] Hoefer, (2002), pp. 210

context and background conditions in which the match was struck that we have described a goodly region of that time slice. Hoefer invokes this notion but runs it in reverse: my free act at a time slice entails nothing about the way the past time slice is outside of those conditions already specified in describing the act. Hoefer puts this point forcefully by describing the regions which a free act completely determines in terms of light cones. I discuss this only to show exactly how trivial the consequences he is attributing to our actions are – relativistic notions such as light cones are not part of his argument, and play only an explanatory role. Imagine a free act and its context: my sitting in my chair looking at a room and deciding to raise my hand. This act entails that one of a large number of microstates obtains. If we suppose that this act and context cover an area of 10m then 'this microstate … determines the past and future microstates only over past- and future-pointing light cones which exist for an absurdly short period of time.'[41] The 'absurdly short period of time' is the time it takes for light to travel 10m. After this period of time, our acts can be pre-empted or undermined and the world can variously interfere with us. Combined with the claim that there are countless microstates that can serve as supervenience bases for macrophysical events we then have a picture where all that is required is that there is some set of deterministically related microphysical states that 'fit' (in accordance with the laws) with the macrophysical history of the world – which, Hoefer thinks, shouldn't be a problem given that there are so many and they obtain for such a vanishingly small amount of time.[42]

I think that there are numerous worries about the account above. I think it is likely that, if Hoefer's account seemed plausible, one would be justified in being deeply suspicious of the technicalities involved; it is firstly not entirely clear how bi-directional determinism is supposed to ground downwards causation – causation running macro to micro rather than vice versa. Presumably a defence of this claim would turn upon Hoefer's insight that the past does not uniquely determine the present. Most of the concerns about downwards causation revolve around the epiphenomenal nature of the higher order properties involved; since there is *already* a complete explanation given of those properties in terms of prior facts about the lower order. Hoefer's B-series determinism allows us to consider the present fundamental in some sense, and so presumably such worries about epiphenomena vanish. I am not sure this is the case. Second, it may seem as though Hoefer has introduced something like *in*determinism into the picture given the fact that there are very many microstates that will

[41] Hoefer, (2002), pp. 211

[42] Hoefer, (2002), pp. 214. At this point, Hoefer is discussing worries about the possible 'harmonization' of these microstates – and so has taken the above account to be cogent. I will argue it is not.

suffice to ground single macro-histories. Hoefer's response would of course be to press the divide between micro and macro and the claim that he has not smuggled indeterminism into his account because it is not the case that some microstate is good enough for *two* macrostates – it is not the case that the very same microstate could ground two macrostates – but at this point I think the argument would devolve into qualms about exactly how fine grained terms like 'good enough' are supposed to be on supervenience. Luckily, I do not think that it is necessary to have these arguments. While the technicalities of Hoefer's account may harbour problems we need not descend into these technicalities to ground strong objections. For the essential character of his defence has already emerged: the triviality of constraint – and this is problematic.

It is the symmetry of the vanishingly weak determinations that our acts come out as having on Hoefer's account that should ground our first worry. For the above has bought us room to claim that our acts do not have consequences towards the past at the price of the same being true towards the future. But our acts do have consequences towards the future. To invoke an example of David Lewis': if Nixon had pressed the button, there would have been a nuclear holocaust. That is a *serious* consequence – having serious effects for the surface of the earth and its future, a non-trivial region of space-time. Not all acts have such serious consequences but they all have consequences beyond the 'trivial' attributed to them by Hoefer's theory. The freedom to determine some tiny region for fleeting amounts of time is not the kind of freedom that is at stake in the debate about free will and determinism.[43] Hoefer *must* also allow our acts to have macrophysical and comparatively long-term consequences for the future – that is a fundamental feature of acts.

We must be careful here in drawing out exactly what is at stake in this objection, for at least part of Hoefer's account above is correct. My choice to act *is* only a tiny portion of a vast time slice. What Hoefer is trying to address is the incompatibilist charge that, if an agent is to be free, she can somehow *choose* to determine large portions of the time slices to the past. This is not the case, but the same is true of the future. Nixon's choice to press the button, does not determine, by that very choice, the time slices that contain the vast irradiated wasteland of earth's nuclear winter – there are various ways such a state of affairs could fail to obtain. The world needs to co-operate with us – and it usually does, at least towards the future. It is the whole time slice in which Nixon presses the button that completely

[43] Brennan, (2007), pp. 211

determines the subsequent state of the earth. Our choices *are* trivial portions of a time slice but the consequences that our choices typically have (in the absence of defeaters and so on) are not. And Hoefer's problem is that it is certain that our choices do have consequences which go towards the future. So why is the same not true of the past, given symmetry? Now, since there is such a focus on symmetry in Hoefer's theory, it seems natural to expect him to account for this fact by invoking a similar sort of macroscopic consequence relation to the past. But this is precisely what he is trying to avoid doing because granting such consequences to the past plays into the hands of incompatibilists who claim that his account then involves unacceptable backwards causation. For Hoefer, at this point in his theory, it would not be backwards 'causation', but it would still be backwards 'consequence' of a non-trivial and macroscopic nature.

Recall the three potential worries that I highlighted in Hoefer's brief statement claiming his thesis did not involve such backward causation:

(i) he invokes the notion of constraint rather than causal efficacy;
(ii) he claims such constraints are weak, given they entail only that microphysical states obtain and;
(iii) that such constraints are *especially* weak towards the past – there is an asymmetry.

In committing himself to (i) and (ii) he indeed deals with the incompatiiblist's charge of backwards causation. Free acts do not cause the past they only place constraints upon what the past is like. But when we see the rest of his account, it turns out that such constraints are both weak and symmetrical – which seems to miss something about the way our acts relate to the future. Basically, at the cost of claiming no backwards consequence he seems to have essentially committed himself to no forwards consequence either, merely these rather weak logical constraints. He recognizes the problem[44] and (iii) is an attempt to deal with it. But (iii), the commitment to re-instating a real asymmetry into his account, seems to undermine freedom from the inside out. For his basic claim is that the challenge to free will from determinism has come from arguing about determinism from an A-series perspective – which is asymmetrical but which asymmetry can be revealed as suspect by invoking the B-series.

Hoefer supports (iii) by suggesting that there are 'pervasive and fortunate' circumstances that ground our inability to have causal consequences towards the past. These

[44] Hoefer, (2002), pp. 212

circumstances are the temporal asymmetries given by the second law of thermodynamics as well as various other unidirectional physical phenomena such as the radiative and fork asymmetries.[45] Hoefer presents an example to the effect that such asymmetries ground our inability to have causal consequences towards the past. If one were to place ice into a glass of water, we never need to worry about the ice getting larger and the water hotter.[46] The way the world is, thermodynamically, precludes this effect. But, Hoefer claims, if this law were to switch temporal directions every second day then it would be the case that we could *cause* a hot drink to have been present *in the past* by adding ice to it on the alternate days. But, the contingent fact is that things are the way they are, and so we cannot do this.

It is important that this asymmetry is not thought to be implied by physical laws, in such a case, the asymmetry of certain physical phenomena would be a feature of determinism: '[Thermodynamic asymmetry] is held to be one that just happens to hold in the world as we know it. It is held to be a *de facto* asymmetry, rather than one required by the laws of physics.'[47] Basically, the world only 'co-operates' with our actions towards the future – but this is not due to determinism or physical laws. And this, for Hoefer, explains our inability to have causal effects over the past but only towards the future – while keeping the issue distinct from determinism.[48] Causal concepts depend, for Hoefer, on 'fortunate' physical asymmetries.

I do not wish to seriously tackle whether or not this is the case – while I appreciate the example, I do not know nearly enough about thermodynamics to be able to confidently claim that Hoefer is right or wrong about this. In particular, would such a case not amount to a reversal of the arrow of time and the A-series? I'm concerned that it might but not sure how

[45] Thermodynamic asymmetries essentially claim that the entropy of closed systems never decreases, it either remains constant or increases – this therefore grounds an increase in entropy that is past-future only. The radiative and fork asymmetries both attempt to give formal explanations of the observation that truly robust co-incidences very rarely occur. Imagine a case in which a still pool of water is disturbed. In such cases, we observe that the resultant wave fronts have various correlated features. They move at a similar speed, they move away from the point of initial disturbance etc. In such cases, the co-incidence of properties possessed by the distinct wave fronts is explicable in terms of a common event in their past. The same is never true towards the future. We never observe 'inwardly moving spherical wave fronts converging to a point' (Popper, 1956) Such asymmetries are thought to be irreducible to the basic, time-reverse invariant laws of physics, although some have tried to so reduce them into probabilities with mixed success. (Horwich, 1987, pp. 60-72). The idea is then that such asymmetries are something over and above what classical physics can describe with deterministic laws. For a more complete account of these asymmetries see both Horwich (1987) and Price (1996).

[46] Hoefer, (2002), pp. 212

[47] Price, (1996), pp. 138

[48] Hoefer, (2002), pp. 212-213

to press this point without a serious discussion of thermodynamics. But I do want to point out the kind of strange argumentative space in which such a claim places him, and to suggest that it undermines his theory. It seems to me that a key part of Hoefer's project is to place a serious divide between determination and causation: between the symmetry of basic physical laws and the asymmetry of causal efficacy. This is not by itself incorrect. But on Hoefer's theory it has the result of our actions being trivial towards both the future and the past because he insists on merely trivial determinations. His response to this is to save causation by basing it upon temporally asymmetric phenomena which are over and above the bi-directional deterministic laws with which he has so far been dealing. This also is perfectly acceptable. It is in fact what many philosophers interested in the arrow of time have been doing for quite some time. So-called 'Third Arrow' strategies attempt to analyse causation in terms of the thermodynamic, radiative and – especially - fork asymmetries, with varying degrees of success.[49] But in attempting to save causation and ground it in physically asymmetrical processes that are *real*, not a result of mistaken A-series type thinking, Hoefer seems to both undermine his own thesis and present the perfect objection to the incompatibilist. Surely it is *this* asymmetry, the asymmetry in causation based upon phenomena taken to be irreducible, that then endangers our free will? If it does not, then Hoefer owes us an argument. He does not give one. It is the lack of such an argument that leads Brennan to conclude that "Freedom from the Inside Out" is incomplete.

3.5 A potential Kantian solution

Can Hoefer's theory be completed? Given the above, it seems he may choose between a couple of possible revisions to his theory. The first is he may simply bite the bullet, remain committed to all of (i-iii) – provide complete accounts of both his theory of 'trivial' action and arguments suggesting how his Third Arrow analysis of causation does not provide a potential re-formulation of incompatibilist concerns founded upon this asymmetry. I have no idea what such an argument would look like if it still had to make use of Hoefer's original thesis. This is a demotivation worry – if Hoefer defends his theory this way then it seems that the A-series and B-series play no part in his argument. All the philosophical work to be done turns upon a discussion of whether physical asymmetries ground a notion of causality that is problematic for our ability to do otherwise. So Hoefer's original argument becomes beside

[49] Price, (1996), Chapter 6; Horwich, (1987), Chapters 3 & 4

the point. Further, such an effort seems mysterious: *any* physical asymmetry that can ground causation seems to provide incompatibilists with exactly what they want – a claim that there is something about the past that does work.

As such, my discussion of alternatives will focus upon giving up some member or members of (i), (ii) and (iii). The most obvious is that he should give up his commitment to (ii) – the weakness of constraints and the strange, quasi-indeterministic (I think) microphysical assumptions that ground it. It is this commitment, regardless of the correctness of the microphysical theory, that results in actions being *trivial*. So it seems likely that (ii) should be rejected. In so revising his theory he may or may not remain committed to (i) – the distinction between constraint and causation. It is the conjunction of (i) and (ii) that necessitate the problematic move of (iii) – for (i) and (ii) entail that acts are symmetrically trivial and this seems wrong. If he gives up on (i) then he must formulate a theory of freedom based upon backwards causation. Such accounts are possible; Peter Forrest[50], in particular offers a theory involving 'backwards causation in defence of free will' in which actions have serious macroscopic causal consequences for both the past and the future. Brennan also suggests that this is what Hoefer should do.[51] I do not particularly like this solution, not only because I suspect it may be fraught with problems. It is obvious, given many incompatibilist arguments against free will, that an account which grants backwards causation can meet them. For their arguments just are arguments to the effect that compatibilism requires backwards causation. Most fundamentally however, I do not like how modifying Hoefer's thesis in this way makes his original insight – regarding the A- and B-series and symmetries in determination – completely besides the point. This is again the demotivation worry. Theories defending backwards causation will have to say something about the metaphysics of time but they will not be based *upon* a distinction between the 'two times'. As I have stated above, I think Hoefer's insistence upon a metaphysics of time itself providing a possible argument is interesting. As such, I would like to provide an alternative for Hoefer that retreats to his original, good, insight: the distinction between the A- and B-series. I suggest that Hoefer commits himself to (i), gives up on (ii) and explains causal asymmetry in terms of the asymmetry of the A-series – in terms of an asymmetry in *agents* as they experience the world.

[50] Forrest, (1985)
[51] Brennan, (2007), pp. 214-215

I will provide only the briefest sketch of such a position on causation – how it works and what features of the account may seem immediately objectionable. A full discussion and defence of what I will call 'perspectival' causation (following Huw Price) would lead this paper much too far afield from its central concern: free will and the free will debate.

The basic notion, which clearly has its roots in the kind of analysis Kant gave of the Copernican revolution,[52] is to claim that the asymmetry of causation is not an asymmetry grounded in the physical world but one which arises out of an asymmetrical feature of agents themselves. There are various agential asymmetries one may point to. The knowledge asymmetry couches the fact that we remember the past but not the future. The deliberative asymmetry appeals to the notion that we act for the future but not the past. The deliberative asymmetry was first (at least after Kant) appealed to by Frank Ramsey[53] in order to explain causation in terms of means and ends. We naturally think of causes as those aspects of reality that we could manipulate – as means – in order to achieve some desired end – or effect. Given some events A and B, we analyse the causal relation by saying that, if A is the cause of B, then A is a potential means of bringing about B. In so doing we locate and explain causation in terms of the kinds of views agents must take upon the world if they are to be able to act. The point is that deliberation is a process that is deeply asymmetrical and fundamentally tied to an A-series view of the world – a world which naturally gives rise to causal thinking. And so the causal asymmetry is attributed to the asymmetry of the experience of agents. I enjoy the way Huw Price puts this point: 'The reason we see asymmetry everywhere we look is that we are always looking through an asymmetric lens.'[54]

Of course, a huge amount of work is needed. In particular, one has to defend against the most obvious objection: that the above account is circular or question begging regarding causation. Just one instance of this sort of objection: it could be argued that the reason I know more about the past than the future is that past facts cause me to have various sensations which cause me to have various beliefs which cause me to act in certain ways and so on. For it could be the case that, were causation to be a feature of the world, not of agents, then it would be *this* fact that would ground the asymmetry of agents and not the other way

[52] Kant, in Guyer (2007), Preface to the second edition of the *Critique of Pure Reason*, B xxii: "Failing of satisfactory progress in explaining the movements of the heavenly bodies on the supposition that they all revolved round the spectator, he tried whether he might not have better success if he made the spectator to revolve and the stars to remain at rest."
[53] Price, (1996), pp. 17
[54] Price, (1996), pp. 158

around. It is around this question that much of the philosophical back and forth regarding the position of perspectival causation has taken place. I think that there is a reasonable case to be made for perspectival causation – that it may be shown that there is some feature of the asymmetry of agents that cannot be claimed to be merely a function of an asymmetry in the world.[55]

The real asymmetries in the agential perspective may have to do with deliberation and practical thinking. Part of the intuition has to do with us knowing more about the past than the future but there is a vital further part. It is that we are *in* the world about which we can know things that renders the agential perspective unique. There is a degeneracy of our knowledge about the world that turns not upon the world but upon the fact that we are part of the reason the world is the way it is. We think of the past as closed and the future as open because the future is open from our perspective – this is a *necessary* perspective (in the sense that we cannot give it up) and grounds an asymmetry in events regardless of any real asymmetry that obtains. We think of events in the past and events in the future as different regardless of whether this is so – we would think this way if physical asymmetries obtained or not. The future takes on the character of being different from the past because we cannot know about it. Not only due to a knowledge asymmetry that may be explained by real physical asymmetries, but because part of what we would have to know is what it is that we do. But the entire process of agential deliberation is directed towards *deciding* what to do. The agential process is degenerative in the sense that the only facts it can countenance are facts that have to be considered as different from other facts – as asymmetrical. In deliberating we must think in terms of a difference between certain facts as they are located in time, we must think causally.

Consider David Lewis' theory of counterfactuals. Counterfactuals are integral to our deliberative process. A large portion of the logic of deliberation is attempting to decide what would happen were I to do x and evaluating the comparative desirability of that outcome relative to others. Counterfactuals are also used by Lewis in an attempt to provide an account of causation[56]. I discuss the following only to briefly sketch what I think is key to the theory of perspectival causation: that deliberative and causal concepts very closely track one

[55] See Huw Price (1996; Chapters 6 & 7) and Jenann Ismael (2007) for two detailed accounts defending perspectival causation. Ismael's account is particularly intriguing in its invocation of an epistemic degeneracy that arises in agents given that they represent a system (the world) as involving their own acts and therefore necessarily asymmetrical and causal, outside of any concerns about a physical basis for such causation.
[56] Lewis, (1973); (1979)

another. There is an implicit but powerful asymmetrical logic to our counterfactual thinking – an asymmetry that I take it Lewis thought was similar enough to causal asymmetry that it could be used to provide a potential means of analysing it. But the counterfactual asymmetry is at least partly a function of the way we use counterfactuals. Here is an example. Imagine I am in possession of a table which has, near either end, two panels perpendicular to the table which may be rotated so as to present different angles of collision for marbles that I roll along this table. I roll marbles from the right side of the table; the panel closest to this side – the panel that marbles will hit first – we shall call P. The other panel we shall call Q. Suppose I have rolled a marble and it has hit a P and is heading towards Q. It seems natural to accept the following counterfactual: If P had had a different orientation then the marble would have a different course. Part of the logic of this counterfactual is that it tells us what we *should have done* if we wished the marble to have a different orientation. The fact comes out as having a special kind of status: it is a potential point of intervention, and it retains this status regardless of whether the laws are asymmetrical or symmetrical. The fact has a special status as causal and asymmetrical for *us*. This turns out to be important. Consider the following 'back tracking' counterfactual: If Q had had a different orientation, then the marble would have a different course. This counterfactual strikes us as false, but let us examine it more closely. The first counterfactual involves holding fixed the past up until the intervention – that is, we hold fixed the past up until the wall having a different angle relative to the marble. The second counterfactual however, when we apply the same logic is *obviously* false, even trivially false in fact. For if we apply the same counterfactual reasoning what we do is hold the past fixed up until the intervention, that is, the moving of the panel, but then it is obviously the case that the marble would not have a different course. But this is because we have *included* this fact in our stipulation of the counterfactual situation. If we could know how the marble would be travelling after its collision with Q we might think differently – that is, if we could hold the future fixed and the laws were bi-directional. But we cannot know this because we hold that fact open, it is what we are trying to decide about. That is, it is part of how we use counterfactuals deliberatively that leads us to use them to ground an asymmetry in facts. There is, in a very real sense, no deliberative reasoning that we can use to figure out what we should do to affect the past, because there is, in a very real sense, no way to deliberate about our actions without holding the past fixed. So, regardless of deep physical asymmetries, a type of causal thinking *must* arise. But this is a feature of our agential perspective.

Despite this woefully brief sketch, I think that perspectivalism about causation provides the kind of account that Hoefer needs if he is not to abandon the original motivation for his project. I am not committed to its tenability and I have not made even the briefest attempt to defend it – but I hope that its general character has emerged. It can *explain away* causal asymmetries in virtue of the one asymmetry Hoefer does have at his disposal, the A-series, the temporal ordering that fundamentally depends upon agents who live in a phenomenological *now*. Hoefer notes the Kantian character of his main thesis[57], the distinction between the A-series temporal world that appears to us and the static, B-series world of physics and determinism.[58] When really pressed on the seeming asymmetry of action, he falls back upon physical asymmetries; he gives up the Kantian point and so undermines his entire project. I think he should insist upon it. He should make another Kantian move. He should claim that the eruption of causation and asymmetry has to do with some form of idealism - the fact that *we* are in the A-series and this has nothing to do with other underlying B-series facts about physical asymmetries that themselves do some work (this would of course require an argument against 'Third Arrow' strategies – attempts to ground causation, predominantly, on fork asymmetries). And we should point out that this is not *ad hoc* in the sense that there are A-series asymmetries which are real given *only* bi-directional physics and no strict physical causation. Namely, A-series embedded agents and the way this affects such agents' counterfactual reasoning and gives rise to a degeneracy in the deliberative stance that grounds a necessary asymmetry. It is worth noting that such perspectivalism is not *eliminativism* about causes. They are real, in the same way that, for instance, secondary properties like colour and taste are. They are there but they are not aspects of the world that physics investigates; they arise partly due to our own particular structure and subjective experience. On perspectivalism, the same is true of causation. It is not an *error*, but it is also no part of the domain of physics. I think that, upon first encountering such a theory, many are likely to baulk, believing that giving up on serious causation is too high a price to pay. This may be the case but it will depend on arguments about causation.

It is this fact - that causation seems to play a key role in *resolving* what has been presented as an incompatibility between only free will and determinism - that I will focus

[57] Hoefer, (2002), pp. 219

[58] It is, in fact, remarkable how similar some of the notions are – the thesis of Hoefer seems neatly summed up by Kant's comment that 'Reason … is not itself in time and does not fall into any state in which it was not *before …* it is *determining*, not *determinable*.' (my emphasis). Kant, *Critique*, A556/B584

upon next. Brennan thinks that if Hoefer's account (and perhaps all compatibilist accounts) will have to supply a theory of causation to meet the incompatibilist's claims then this renders the theory irrelevant in so far as it fails to take a position on causation. This is not the case – this does not render such theories irrelevant as they relate to highlighting features of the *debate* about freedom and determinism. For I argue the incompatibilists need to provide an account of causation too – the burden is not only upon their opponents. I think that the symmetry of the B-series and determinism and the first half of Hoefer's thesis have serious lessons to teach us about incompatibilism as a thesis. Namely, that incompatibilists have been labouring under an assumption about asymmetries that has led them to argue without invoking causation. And their failure to make this assumption explicit – and to argue for it – has resulted in serious gaps in the formulations of their arguments. My essential thesis is the following: the metaphysics of time given by the B-series and the block universe reveal failures in both classical incompatibilist arguments and the consequence argument. Hoefer believes that this creates the room to mount a *defence* of freedom but I think that this is too strong. The defence of freedom must require a long and involved theory of causation. I think that Hoefer's insight can only be properly deployed in defence of a much weaker thesis: that incompatibilist arguments are not so strong as they may first appear and that the entire debate between free will and determinism has been slightly off the point. I will develop these claims in the next section.

4 Asymmetries, Causation and Costs

Let us briefly recap: incompatibilism claims that the free will thesis and the thesis of determinism are not co-tenable; they cannot both be true. What the block universe reveals, and what Hoefer attempts to make use of, is that mere determinism, as defined by incompatibilists in their arguments, is not *enough* to ground an incompatibility with the free will thesis. For Hoefer's essential insight reveals that there is no reason to take the past as fundamental in the relations of determinism and so it is not the case that the past does work in 'closing' off the alternative possibilities of free agents. Relations of determination according to deterministic laws run from future → past as well as past → future. Such a perspective creates the kind a kind of negative room for freedom: determinism does not *rule it out*. But, as Hoefer's positive thesis shows, there is a rather steep price to pay for taking up such a perspective. One must give up on causation. For the determination of the past by my free actions in the middle of the block universe is not a causal relation. But it is symmetric. And so it is not the case that the relation between my free actions in the middle of the block universe and *future* time slices is causal. Because of Hoefer's peculiar account of action, this turns out to have the further unfortunate consequence of actions being trivial. In trying to deal with this Hoefer makes a mistake and re-introduces causation – or at least some sort of physically grounded, temporally asymmetric relation – and provides himself with a dilemma. Either this new asymmetry is a threat to free will or it is not. He does not provide an argument to the effect that it is not, but I have sketched what such an argument would look like. It would be a denial of causation 'out there' in the world – explaining the phenomena away in terms of agential asymmetries. We must admit the cost of this: no 'real' causation.

I think that at this point, many incompatibilists and champions of the Consequence Argument would feel vindicated. 'Robust' compatibilism of the sort advanced by Hoefer – that is, compatibilism that takes free will to consist in the ability to do otherwise – seems unworkable. But this is too hasty. In light of Hoefer's first insight and the resultant claim that determinism *doesn't* preclude free will, we need to revisit the Consequence Argument and make sure this is correct. Either there is something wrong with the Consequence Argument or there is something wrong with Hoefer's *core* claim and his entire paper is a failure. I think there is something wrong with the Consequence Argument – without a particular assumption about the nature of time it does not truly establish incompatibilism.

4.1 Two versions of the Consequence Argument

Recall Van Inwagen's statement of the basic Consequence Argument:

> If determinism is true, then our acts are the consequence of laws of nature and events in the remote past. But it's not up to us what went on before we were born, and neither is it up to us what the laws of nature are. Therefore, the consequences of these things (including our present acts) are not up to us.[59]

In *An Essay on Free Will*, Van Inwagen presents three detailed versions of this argument which support the same conclusion – that the free will thesis cannot be true at deterministic worlds – and which stand and fall together.[60] I will present the first and third of these arguments and present some objections which I submit undermine the strength of their conclusions.

For both arguments, the formal definition of determinism is given by the conjunction of the following theses:

> (D1) For every instant of time, there is a proposition that expresses the state of the world at that instant;
>
> (D2) If *p* and *q* are any propositions that express the state of the world at some instants, then the conjunction of *p* with the laws of nature entails *q*.[61]

4.1.1 The First Consequence Argument (FCA)

The **FCA** asks us to imagine a case where some individual, J, refrained from raising his hand at some time t. P is a proposition that expresses the state of the world at t – at least part of what P expresses is that J did not raise his hand at t. P_0 is a proposition that expresses the state of the world at t_0 – a time in the remote past. L is a proposition that expresses the deterministic laws of the world. We then construct the following argument.[62]

> (F1) On the assumption that determinism is true, the conjunction of P_0 and L entails P.

[59] Van Inwagen, (1983), pp. 56
[60] Van Inwagen, (1983), pp. 56-58
[61] Van Inwagen, (1983), pp. 65. This is, of course, equivalent to my definition of determinism given in the second section of this essay.
[62] Van Inwagen, (1983), pp. 70

(F2) It is not possible that J have raised his hand at t and P be true.

(F3) If J could have raised his hand at t, then J could have rendered P false.

(F4) If J could have rendered P false, and if the conjunction of P_0 and L entails P, then J could have rendered the conjunction of P_0 and L false.

(F5) If J could have rendered the conjunction of P_0 and L false, then J could have rendered L false.

(F6) J could not have rendered L false.

(F7) Therefore, if determinism is true, J could not have raised his hand at t.

If J could not have raised his hand at t, then it seems as if J was not free at t. For free will requires that, while he exercised his ability to refrain from raising his hand, he could have exercised the ability to raise it. Like all versions of the Consequence Argument, the above turns upon pointing out a kind of necessity that the past and the laws have, at least as regarding our relation with them. In this case, agents 'cannot render false' propositions expressing past states or propositions expressing the laws of nature. I think that most of the premises of the above argument are reasonably self-explanatory – although I need to make some comments about (F4) and (F5).

(F4) seems to rest upon a principle that I will call closure.

Closure: If an agent S can render Q false, and P entails Q, then S can render P false.

Van Inwagen thinks that such a principle is a 'trivial truth'[63] and it certainly does seem that way as long as we take the ability 'to render false' as logical – that is, if closure is understood as expressing a closure under *entailment*. That is, it is a trivial truth if it is based upon this fact: if P entails Q, then the denial of Q entails the denial of P. Concerns about exactly how the ability 'to render false' should be understood will be important in my discussions to come.

(F5) is not particularly controversial. If I can render a conjunction false, presumably this is because I can render one of the conjuncts false. What is intriguing is that Van Inwagen chooses the particular conjunct he does – he chooses L. Van Inwagen chooses this conjunct

[63] Van Inwagen, (1983), pp. 72

because he is committed to the principle that one cannot change the past – and so *obviously* nobody could have the ability to render false a proposition about the past. Concerns about this principle will also form part of my discussion of this argument. For now, it is just necessary to point out that the exact same argument can be made by replacing the instances of 'L' in the antecedent of (F5) and in (F6) with instances of the historical proposition 'P_0'. If Van Inwagen is right about his general principle, this may even strengthen the argument.

4.1.2 The Third Consequence Argument (TCA)

TCA makes use of a modal operator 'N' ranging over sentences *p*. N*p* means '*p* and nobody has, or ever had, any choice about whether *p*'.[64] He then provides us with the logic of this operator.[65]

(α) from □*p*, deduce N*p*

(β) from N(*p* → *q*) and N*p*, deduce N*q*

Assuming that 'P_0' and 'L' denote the same propositions as they do in the **FCA** and P is a true proposition expressing the state of the world, we then construct the following argument:

(T1) $\Box((P_0 \wedge L) \rightarrow P)$	determinism
(T2) $\Box(P_0 \rightarrow (L \rightarrow P))$	from (T1) by exportation
(T3) $N(P_0 \rightarrow (L \rightarrow P))$	from (T2) by (α)
(T4) NP_0	premiss
(T5) $N(L \rightarrow P)$	from (T3) and (T4) by (β)
(T6) NL	premiss
(T7) Therefore, NP	from (T5) and (T6) by (β)

According to **TCA**, we never have any choice whatsoever about any true proposition. But if we are to be free then, in choosing which of our abilities with respect to a contemplated act we will exercise, we must have a choice about P. So **TCA** seems to conclude to

[64] Van Inwagen, (1983), pp. 93
[65] Van Inwagen, (1983), pp. 94

incompatibilism. I will examine one putative objection to this argument before making my comments.

4.2 Consequences and assumptions

In 'Free Will and the Necessity of the Past'[66], Joseph Campbell presents an objection to the above argument. The sketch of his objection is as follows. He identifies two possible lines of support for one of the premises in **TCA** and rejects one these lines. He then claims that in taking the only tenable line of support, the proponent of **TCA** must make use of a further contingent proposition: the existence of a remote past. Campbell thinks that such a move is problematic for **TCA** because it renders incompatibilism a contingent thesis whereas it is supposed to show that at *no* deterministic world does the free will thesis obtain. I grant Campbell's argument but think that there are slightly different conclusions to be drawn from it.

Campbell wonders about the support for (T4), the premiss to the effect that NP_0. He examines the two possible reasons for holding NP_0 discussed in *An Essay on Free Will*.[67] Either, as Van Inwagen notes, we hold NP_0 because "P_0 is a proposition about the past and "nobody can change the past."' Or, we hold NP_0 because 'P_0 is a proposition about the "remote past."'[68] That is, either NP_0 because of the pastness of P_0, or NP_0 because of the remoteness of P_0. But worries about pastness cannot ground NP_0. For imagine a drunk driver who is sober at t1, drinks at t2, and then kills some individual A, at t3. Such a drunk driver cannot appeal to 'N' at t3. For N states that nobody has and nobody *ever had* a choice about some proposition. But while it is true that at t3, the drunk driver did not have a choice about killing A, it seems wrong to say he *never had a choice* about this fact. For he seems to have had just such a choice at t2, when he began drinking. So, the worries about P_0 cannot be merely due to the pastness of P_0 – something else is at stake. Campbell thinks that it is the *remoteness* of the past that does the work: 'If determinism is a threat to our free will, it is only so because there are true propositions about the remote past.'[69] Campbell then presents a counterexample to this requirement. There is a possible deterministic world at which there

[66] Campbell, (2007)
[67] Van Inwagen (1983), pp. 92 and 96
[68] Campbell, (2007), pp. 106
[69] Campbell, (2007), pp. 109

has, from the first moment of this world's history, existed an individual 'Adam'. For Adam, there is no proposition about the history of that world which will satisfy N – for he has existed at every instant of the world's history and so, at some time, it will be false that he *never had a choice* about that proposition. But then there exists a deterministic world at which the Consequence Argument does not seem to establish the non-existence of free will. Campbell thus concludes the Consequence Argument '[does not] offer general reasons for endorsing incompatibilism.'[70] For incomatibilism to really be established, the argument needs a contingent premiss that, for every individual agent, there exists a remote past such that the agent did not exist then.

Campbell's objection has some merit, but I think it needs quite a bit more clarification. I am struck by the similarity between Campbell's argument and Hoefer's brief statement to the effect that the deterministic challenge to free will has something to do with an assumption on the part of incompatibilists. Hoefer thinks that assumption is about time. Campbell seems to agree that the assumption is about time but – recalling the B-series – there is something strange about his concluding that the assumption turns upon a time's *remoteness*. For, on the B-series, if (as Campbell argues) *pastness* does not do any work, why should a time's being *very past* do any? If a time's being past grounds nothing, how can a time's being 'more' past ground something? The answer, of course, turns upon the kind of agents we are. A time's being very past, relative to *us*, does do something – but it doesn't do anything to that time. It does something to us. We, intuitively, lose an ability when we consider times that are not just in the past but in the remote past. We lose the ability to have had causal control over the facts of that time. We think, when we consider times in our past – like the case of the drunk driver – that while we do not have a choice *now* about what happened at that time we did have a choice *then*. But when we consider truly remote times, we do not think that we *ever* had a choice. So Campbell is broadly correct. As it stands, the Consequence Argument seems to require a contingent premiss to the effect that a remote time exists. But when we understand what is really at stake in *remoteness* – control – we see that all the Consequence Argument really requires is that there be a time beyond our control. For Campbell, this amounts to a remote time, but it is *control* that is at stake. I think we are getting near the essential problem with the Consequence Argument. It is an argument about determinism. On the Consequence Argument, determinism is supposed to *result in* an

[70] Campbell, (2007), pp. 110

absence of control but it only seems to do this when *combined with* such an assumption.[71] Determinism itself is not enough.

At this point I think it would be fair to grant that something strange is going on. For when we see that Campbell's reply truly turns upon a notion of control one might worry that Campbell has missed a point. For the **TCA** functions with 'N' – an operator to the effect that we do not have control. The problem is that N stipulates a lack of control which may require an argument. I will draw this out in a couple of ways.

First, let us return to the **FCA** – in which our individual J does not raise his hand at t and so on. Given Van Inwagen's definition of determinism, this argument will remain perfectly valid if we run it in the opposite temporal direction; that is, if we run it future → past. Let us do so. If we run this argument in the opposite direction then there is some proposition P_1 about the state of the world at time t_1 which is in the remote future relative to t. P is a proposition about the state of the world at t and at least partly expresses the fact that J does not raise his hand at t. L is a proposition expressing the laws of nature. P_1 and L jointly entail P. Now, if J can raise his hand at t, then J can render P false. But P_1 and L jointly entail P, so, if J can render P false then he can render false the conjunction of P_1 and L. Sticking to Van Inwagen's formulation: if J can render the conjunction of P_1 and L false, then J can render L false (this following from the mirror image of Van Inwagen's general principle: one cannot change the future). J cannot render L false. So J cannot raise his hand at t. Q.E.D.

This is a bad argument. But it is the *same* argument – at least as it relates to determinism. Why is this a bad argument? Perhaps it is because one disagrees with the 'general principle' here – that one cannot change the future. But the ability to *change* the future is surely a very queer ability. To change the future – just as in the case of changing the past – is to be able to make some future fact be some way at a time and another way at another time. This seems incoherent. Much better – perhaps - is to say 'affect the future' – or perhaps, most intuitively, to make the future to be the way it is. This is, of course, what the incompatibilist who advances the Consequence Argument believes. There is a difference between the future and the past – we can affect the future but not the past. And, most precisely, we can make the future be the way it is – but not the past. But *the argument does not ground this*. Nothing about determinism grounds this ability. If only determinism were at stake, then the above argument would be a good one – it would *result in* our lack of control.

[71] Sartorio, (2014), pp. 262

But it doesn't, in order for the above argument to result in a lack of control, we need a stipulation to the effect that we cannot affect the future – but this needs to be argued for. It is something extra. The same, it seems, might be true on the **FCA** – and its principle. It is worth briefly noting that the above is not supposed to be a serious objection against the Consequence Argument – it is merely meant to show that there is more going on in the Consequence Argument than its premises imply. Determinism doesn't do the work alone.

Here is the same point put in, yet again, a different way. Incompatibilism about determinism and free will is not fatalism about free will. Below is a simple argument for **logical fatalism.**

Where 'A' denotes some action that I perform:

(F1) It is true that I A tomorrow.

(F2) Necessarily, if it is true that I A tomorrow, then I A tomorrow.

(F3) It is not up to me that it is true that I A tomorrow, nor is it up to me that *if* it is true that I A tomorrow then I A tomorrow.

(F4) Therefore, I cannot refrain from A tomorrow.

This argument, too, is a bad argument. It commits a modal fallacy – moving from what is a merely true antecedent, through a necessary material conditional, to the necessity of the consequent.

Modal Fallacy

(MF1) p

(MF2) $\Box(p \rightarrow q)$

(MF3) $\Box q$

Such a pattern of inference is obviously invalid. For example: imagine the following argument. Some individual, J, has a beard. Necessarily, if J has a beard then J has a beard. Therefore J necessarily has a beard. This does not follow. Such an argument erroneously locates necessity, attributing it to one of the terms of a necessary conditional, a term which is merely true. The fatalist confuses truth with necessity and argues that since we have no power

over what is necessary then we have no power over what is true. Now, Van Inwagen is far too fine a philosopher to have made this mistake in his argument. He relies on the following, valid, pattern:

Valid Necessity

Where 'N' expresses Van Inwagen's necessity operator.

(VN1) N*p*

(VN2) N($p \rightarrow q$)

(VN3) N*q*

The problem though, is how Van Inwagen arrives at N*p* where *p* is a proposition about the past. Rule (α) above states: from □*p*, deduce N*p*. But, as the discussion of Campbell has shown, it is not the case that a true proposition about the past is a necessary proposition. That is to think like a fatalist. So N*p* is *not* deduced from □*p*, when *p* is a true proposition about the past. So it is not by (α) that Van Inwagen introduces the premiss N*p*, where *p* is a true proposition about the past. Van Inwagen is not a fatalist, but then *why* N*p*? A serious argument is missing because Van Inwagen assumes it is obvious – we cannot control the past. This point is very similar to the point made in the above discussion of Campbell, it may even be the same point, but it bears repeating. There is an assumption about the 'obviousness' of one of the premises of the Consequence Argument, but the problem is that assumption has something to do with the very fact the argument is trying to establish: a lack of control.

Finally, I will make one last point about a difference between fatalist arguments and arguments from determinism. There is something else wrong with the argument from **logical fatalism** above, besides worries about the modal fallacy. One might take objection with (F3). Perhaps one might claim that it *is* up to me whether it is true that I A tomorrow. It may be claimed that what makes certain tensed propositions true is what goes on at the times that the propositions are about.[72] For example, it seems correct to say that propositions about the year 1500 A.D. are true in virtue of what goes on at 1500 A.D. If this is the case, then it might be that what makes it true that I A tomorrow is that *tomorrow*, I decide to A. It is a timelessly true fact that I A tomorrow, but it is true because I A tomorrow – it is made true by my free

[72] Horwich, (1987), pp. 30-31

choice tomorrow, not some *past* fact *about* tomorrow. This seems to me to be a correct solution to the problem of fatalism. It also seems to me to be the solution to any fatalistic concerns one may have developed while considering the block universe. For on the block universe, there are true propositions expressing what I will do tomorrow. A philosopher in the grip of fatalist reasoning might think that this fact alone means that we cannot be free in the block universe. I think that the above goes some way to defusing this worry. It is in virtue of what goes on at a time that propositions about that time are true or false.

If this is the case then there is another lesson to be learned from comparing the Consequence Argument with fatalistic arguments. For, on the Consequence Argument, there are also true propositions about what I shall do tomorrow. These propositions are entailed by historical propositions and propositions expressing the laws of nature. The same is true in the other direction. The difference between the arguments is that, on the Consequence Argument, there are supposed to be truths about tomorrow that are not true solely *in virtue* of what goes on tomorrow but in virtue of the historical proposition and the laws of nature. Facts about the past and facts about the laws are supposed to be part of the reason that truths about tomorrow are true. But the Consequence Argument does not ground this. Bi-directional entailment and the Consequence Argument does not show this to be the case. *Why* is the true proposition expressing the current state of the world not true only in virtue of what goes on at that time? What *else* is happening? There is an assumption about a particular priority in the direction of determination that is not supported by the argument. But this assumption is doing serious work – it is the *reason* the Consequence Argument is not committing the kind of fatalist mistake discussed above. It needs to be made explicit and argued for.

4.3 The ability to render false

In his paper "Are we Free to Break the Laws?" [73] David Lewis first presented an objection to the **FCA** that, I think, offers the clearest expression of worries similar to the above. For the essential point to be found in the above is that the Consequence Argument does not seem to establish incompatibilism without some sort of assumption about ability and control. Since Lewis, a slew of philosophers have offered similar arguments.[74] The arguments turn upon a close scrutiny of what is involved in Van Inwagen's posited 'ability to render a proposition

[73] Lewis, (1981)
[74] Fischer, (1986); Mackie (2003); Horgan, (1985); there are many others.

false.' Perhaps the analysis of the ability to render false will reveal exactly what kind of assumptions have been made by the incompatibilist – will reveal why it is so 'obvious' that we cannot affect the past by revealing what the incompatibilist thinks such an ability amounts to. I will present the argument only rather briefly and in my own words, for an in depth discussion of the issue would lead us through the entire modern debate between incompatibilism and compatibilism – the circular character of which is easily brought out below.

Essentially, one draws a distinction between a strong and weak reading of the ability to render a proposition false.[75] Let 'S' denote some agent and 'P' denote a proposition. On the strong reading: S can render P false if S can do something such that, were she to do it, P would be false *in virtue of her doing it.* On the strong reading, S can perform a past-altering or law-breaking action: an action that itself causes the past or the laws to be different. On the weak reading: S can render P false if S can do something such that, were she to do it, P would be false. On the weak reading: S's action is not itself a past-altering or law-breaking action, her action does not cause P to be false.

The strong reading finds a natural fit with causation. And it finds a natural fit with the lurking assumption of the champion of the Consequence Argument. For, if we are to take the ability to render false in a strong, causal, sense then, as the incompatibilist claims, the following *is* obviously mistaken. Where 'P_0' denotes a proposition about the remote past: S can render P_0 false in the strong sense. That is, S can do something such that, were she to do it, P_0 would be false in virtue of her doing it. She can perform a 'past-altering' or 'law-breaking' action. This is the sense glossed by Van Inwagen's general principle that one 'cannot change the past' and the weaker principle that one cannot have causal control over the past. Nobody could have the ability to change the past and the ability to have causal control is certainly strange and requires explanation. However, the weak reading claims nothing like the above. The weak reading abuses a counterfactual sense of the 'ability to render false'. On the weak reading, S can render P_0 false just in case S can do something that, were she to do it, P_0 would be false. We might draw this out by saying that, S can do something such that, were she to do it, P_0 would have to have been false. But nothing in the Consequence Argument shows that it *couldn't* have been false. The argument at this point can go two ways. Many philosophers who have advanced this objection and distinguished the

[75] Lewis, (1981), pp. 115

strong from the weak sense then argue that, on the strong reading, the principle of **closure** upon which the **FCA** relies becomes questionable.[76] That is, they maintain that the argument might be invalid. For, **closure** states that: if an agent S can render Q false, and P entails Q, then S can render P false. But the validity of this principle was based upon the logical notion of entailment closure; that a *denial* of Q entails the *denial* of P. This does seem a trivial truth of logic and must be accepted. However, if we are to take 'the ability to render false' in a strong, causal sense then it seems to follow that if S is to render some proposition P false then, S can render P false only if P is true because S has causal control over the propositions from which P follows. But this does not seem such a trivial claim. It certainly doesn't seem to capture how we think about our ability to influence things, we do not think that what is *required* for us to, say, refrain from raising our hand, is that we must have causal control over everything sufficient for our so acting. This is of course what the incompatibilist is claiming – it is the conclusion of the Consequence Argument - but it is no longer such a trivial, logical truth as **closure** seemed to be based upon.

I do not wish to seriously consider this line of attack; it strikes me as equivalent to claims that the incompatibilist is assuming something and it is not the only possible move. For the other way to proceed is to grant the argument's validity and point out that one of its premises seems false - namely the premise that an agent cannot render false a proposition about the past or the laws of nature. On the weak reading of ability, an agent *can* so render a proposition false. Namely, she can do something that, were she to do it, P_0 would be false. But she would not, by virtue of her act, thereby *cause* P_0 to be false.[77] This strikes me as precisely turning upon the acceptance of closure – if closure is valid, in the sense described by Van Inwagen, then the weak reading of ability seems to follow.

This kind of response indeed finally reveals what – in light of the above – we may have begun to suspect: in order the strictly establish incompatibilism, the Consequence Argument must make an assumption about control, ability and physical asymmetry. It must assume a strong reading of the ability to render false, along with a serious sense of causal control. But, the argument doesn't *establish* the correctness of this reading. It seems to follow from the very validity of the argument that we can take the other; it is perfectly cogent to take up the weak sense of 'able to render false'. Again, the Consequence Argument is supposed to show that determinism *gives rise to* a lack of control, an inability to render false propositions

[76] Horgan, (1985), pp. 341-343
[77] Lewis, (1981), pp. 118-121

about the past. But it only truly establishes this to the extent that determinism is *conjoined with* an underlying assumption about such abilities.

At this point, I think the debate reaches a stalemate. There is something circular and question-begging about both the compatibilist and incompatibilist positions. The incompatibilist appeals to the intuitive power of the Consequence Argument: if determinism is true, since we *cannot* render false propositions about the remote past and the laws, we cannot render false propositions about states of world which include our present acts. The compatibilist replies: since we *can* render false propositions about states of the world which include our present acts then, by determinism, we can render false propositions about the remote past and the laws of nature. The incompatibilists replies: this begs the question – what we are arguing about is exactly whether we can render false propositions of the first kind. You owe us an argument. The compatibilist replies: but this begs the question - you have not given us an argument which concludes to the fact that we cannot render false propositions of the former kind except one which claims that we cannot render false propositions of the latter kind. But that is precisely what we are arguing about. You, too, owe us an argument. And so begins a long debate about which is the most fundamental claim and upon whom the burden of proof should lie.[78]

4.4 The costs of compatibilism

I wish to suggest a way out of the vexed mess presented above. I think that Hoefer's insight regarding temporal asymmetries does work here – if only of an explanatory and illustrative nature. In fact, I think "Freedom from the Inside Out" presents us with a microcosm of the modern debate and allows us to take stock of exactly what is at stake. Hoefer claimed that the challenge to free will from determinism resulted from the 'unholy marriage' of determinism with assumptions about asymmetries in time. I think this is correct. In presenting the Consequence Argument, incompatibilists have taken up an A-series perspective upon which temporal and causal asymmetries are *obvious*. However, because the Consequence Argument rests upon such an assumption, it does not truly establish incompatibilism. There are worlds in which determinism and free will are compatible. Campbell's 'Adam' world is just one. But when we see that what is at stake at Campbell's world is control we see that determinism and

[78] See Foley, (1979) versus Van Inwagen, (1980) for just the first in a long line of such replies and counters.

free will are in fact compatible in any world without asymmetrical control. This means giving up on the notion of real causation – and taking up a weak reading of our ability to render propositions false. But such a move is coherent; nothing about the Consequence Argument establishes such facts about control. The Consequence Argument only concludes that free will is incompatible with determinism *and* some further fact. But the further fact is simply assumed. The path remains open, on the Consequence Argument, to take up the weak, non-causal, reading of an agent's ability. The incompatibilists may find this mysterious but they must concede that their argument, proceeding with determinism alone, does not establish their conclusion. Compatibilism remains a live option.

There is a very real cost to such compatibilism however. The 'weak' and 'strong' readings of the ability to render false propositions about the past play the same role as Hoefer's distinction between 'cause' and 'constraint'. In Hoefer's block universe, our actions do not *cause* the past to be the way it is, they merely place *constraints* upon it. On the weak reading of the ability to render a proposition false, our actions do not *cause* the propositions to be false, they merely entail that they are so. Both compatibilists and incompatibilists have lessons to learn here.

Compatibilists who wish to make such a move must grant that they are arguing for an atemporal world. That is, they must accept a B-series metaphysic of time. For on *any* tensed view, their reading does involve (as the incompatibilists maintain) a 'changing of the past.' If propositions which express states of the world in the remote past are actually expressing facts that are *ontologically* prior to later facts (the later facts therefore *dependent on them* if they are dependent on anything), the compatibilist reading cannot work. On the A-series, facts about the past were *thus and so* before any agents existed, before any agents could act freely. One cannot take up the weak reading in such a world. Even if the counterfactual dependence sense of the ability to render a proposition false only entails that, S can do something such that, were she to do it, P would have *had* to have been false - it *wasn't*. The dependence is asymmetrical. The fact expressed by P was there *first* – existed first - and whatever sense in which it would have *had* to be different would require a *change* in what P expressed. This is impossible. Compatibilists therefore have to grant a B-series metaphysic of time if they are to make the weak reading. This is at least partly because counterfactual dependence allows for

'back-trackers' that do not invoke causation.[79] If there is a deep asymmetry in dependence such counterfactuals seem incoherent as descriptions of the world. So compatibilists need to realize that they are giving up on temporally-grounded ontological priority. Incompatibilists, on the other hand, need to grant that this *is* a possible move – compatibilists are not begging the question.

However, compatibilists should not feel too pleased with the space they have carved out for free will. As Hoefer showed, mere bi-directional determinism, without the A-series assumptions, buys us room for free will. For the worry that the *past* determines us disappears. We are equally justified in taking our actions in the present to constrain the past and the future. This is correct, given determinism and a metaphysics on which events, no matter their temporal location, are ontologically equal. The cost, for Hoefer, of this symmetry was its failure to capture what seems to be a natural asymmetry in action. Part of this had to do with actions coming out as *trivial* because of Hoefer's microphysical theory. Compatibilists do not need to commit themselves to *that*, but they need to grant the symmetry of their reading of agential abilities. An agent has the ability to render a proposition false just in case she can do something such that, if she does it, the proposition is false – but it is not *because* of what she does that it is false. This is true towards the future too, on such a reading. This means that the compatibilist must commit herself to saying very strange things about agent's actions as they relate to the future. She must, for instance, say something like the following. Let a considered act be the decision to make a cup of coffee and a proposition 'C' express a state of affairs that includes this cup of coffee: 'I can do something such that, were I to do it, C would be true – but not *because* I did it, in the causal sense.' Exactly as it was with Hoefer, if such compatibilists insist that they are not *causing* the past, then neither are they *causing* the future. And if a compatibilist wishes to take up the weak reading but nevertheless remain committed to a more 'natural' account of agential abilities as they relate to the future then she must argue and explain why this asymmetry does not reveal the incompatibilist reading to be correct. As with Hoefer, perhaps such an argument is possible – but I certainly do not have any idea of what it would look like. I have suggested that the best thing to do would be to

[79] David Lewis, (1973), would of course deny any real back-tracking account of counterfactuals. This is because he is trying to use counterfactuals to give an analysis of causation and so must close such a symmetrical possibility. I have never been particularly convinced by Lewis' account– partly because his theory involves such a baroque treatment of how we should analyse the 'similarity' of possible worlds (local miracles, different laws, but no large scale changes to the world's history) and partly because his 'local miracles' themselves involve tiny historical deviations. Some trivial fact about the immediate past is changed on Lewis' miracle account – this must be the case to ensure that it is not an agent's *action* that comes out as itself being 'law-breaking'.

provide an account of our natural asymmetrical and causal commitments which can explain them away. I think this probably means defending either Perspectivalism or outright Humean eliminativism (although I seriously prefer the former theory).[80] But, most of all, she must admit what her reading entails: no 'real' causation.

The incompatibilists, on the other hand, also face a number of challenges. First, in order to show that the compatibilist reading is outright *wrong* and illogical, they need to argue for a tensed metaphysic of time. Presentism or perhaps one of its more complex theoretical siblings. I have doubts about their ability to do this; arguments for Presentism are controversial at best, especially in light of our current physics. But this is not the only argument available to the incompatibilist. Incompatibilists can argue for causation on the B-series – *real* causation: asymmetrical, physical, necessitation. I think that there are two ways that they can go about this. Either they can pursue an analysis of causation along the lines of 'Third Arrow' accounts. That is, they can ground causation on the *de facto* physical asymmetries of our world that give rise to the arrow of time and do the work of necessitation. There are promising, although disputed, Third Arrow accounts of causation. On the other hand, they can be 'realists' about causation.[81] Such realism, however, seems contentious. Realism about causation, in the block universe, amounts to a kind of metaphysical brute fact. It cannot be grounded in the basic physical laws given by determinism, for they are bi-directional. Nor do the realists about causation wish to pursue Third Arrow strategies – presumably because they think that it is in virtue of causation that these *de facto* physical asymmetries exist. But there is then no physical law or physical asymmetry *left* upon which to base causation. Causation, on the realist view, must be something *over and above* the physical laws and physical asymmetries. There is another metaphysical layer to the world: a causal layer whose structure is not reducible to the physical. Such realism will obviously be contentious.

It is worth stressing that incompatibilists have to countenance these costs. I think that proponents of the Consequence Argument often realize that compatibilists are essentially

[80] Humean eliminativism about causation seems suspect, to me at least, since it suggests that what seems *real*, if only phenomenologically real, is really about words and our stipulated definitions. It seems wrong to say that 'cause' just denotes the temporally earlier term of a constant conjunction. There is more to it than that. Hume's account also seems to rule out backwards causation and time travel by fiat – another contentious move. Perspectivalism has the upper hand here because it attempts to explain away causation without claiming that it is *just* a matter of definition. We can, on Perspectivalism, retain a richer conception of causality than on eliminativism - just so long as we realize that it has to do with us, not the world of physics.

[81] Or, as Huw Price calls them, 'hyper-realists' about causation.

giving up on causation in meeting its challenge. They take this to be an obvious point in their favour, and in favour of the Consequence Argument. But causation is not so obvious a notion; particularly when we reveal that part of the reason that incompatibilists are so convinced by the strength of the Consequence Argument is that they combine determinism with an A-series perspective. On such an A-series perspective, temporal asymmetry and causation are natural and obvious features of the world; it is no wonder incompatibilists did not think they had to include defences of such asymmetries in their arguments. But their arguments present with a *gap* because they have failed to give such a defence. The Consequence Argument does not establish incompatibilism – it does not *rule out* free will, given determinism – until determinism is combined with a further account of asymmetric control. If compatibilist arguments like Hoefer's are incomplete then so is the Consequence Argument.

On the B-series, there is nothing obvious about causation at all. But this doesn't mean incompatibilists are *wrong*. Hoefer is again instructive. Hoefer claimed that the challenge to free will from determinism resulted from the conjunction of determinism with a *mistaken* assumption. What has been revealed is that there is such an assumption in the Consequence Argument – determinism is conjoined with a further commitment to asymmetrical necessitation – but this has not revealed the assumption to be *mistaken.* As I argued, Hoefer thinks that if there is an assumption then this allows a *defence* of freedom. This is too strong; for to show the assumption to be mistaken, Hoefer needs to provide an involved theory of causation (or lack thereof). The same, however, should be the work of incompatibilists. They need to argue for their assumption – for, on the B-series, asymmetrical necessitation is not so easily assumed.

4.5 Determinism, causation and the view from nowhere

A final enquiry remains: has the role of causation in the challenge to free will been missed? The answer is rather obvious, in light of the above. The role of causation has not been missed, it has been assumed. For the basic intuition of the incompatibilists, and the essential idea expressed by the Consequence Argument is the following: if our acts are determined by events beyond our control, then we are not free. This is a persuasive line of thought. The problem is that determinism alone has been offered as the rigorization of this notion.

Consider Van Inwagen's statement about the relationship between determinism and causation: 'I doubt whether universal causation entails determinism. I am uncertain what to say about the question whether determinism entails universal causation … but any real discussion of this question would lead us needlessly, for we need not answer it, into a discussion of causation, something I shall avoid whenever it is possible.' For Van Inwagen, determinism *tout court* grounds his argument. Determinism, on the Consequence Argument, is supposed to couch both the process by which our acts are determined (by physical laws and entailed future states of the world) *and* the fact that they are so determined by a special sort of fact – *past* facts. Facts beyond our control, facts that do controlling. But determinism cannot account for both. An additional explanation of this asymmetry in the status of past and future facts is required.

For those championing the Consequence Argument, determinism alone does the work because they have assumed that there exists an asymmetry in time, in events located in an A-series metaphysic. An assumption in favour of such an ontological asymmetry is natural given our A-series perspective. But it *is* an assumption; and one that is rendered somewhat dubious by the arguments in favour of the tenseless B-series of the block universe.

The block universe is a strange place; almost incomprehensible to fundamentally temporal beings such as ourselves. And if the block universe is at odds with the notion of ontological priority based upon temporal location then it is also very much at odds with any temporally asymmetrical causal thinking. Thomas Nagel coined the phrase 'the view from nowhere'[82] to gloss a truly objective perspective, a perspective outside of time and all the subjective conditions of human experience. From this perspective, the block universe is an unchanging four-dimensional object in which nothing *happens* in the active sense that we observe the world happening around us all the time. Causal concepts struggle in such an environment. The fundamental constituents of this world do not become, or change, or begin, or end – they do not *make* things happen in an active sense; rather the whole of reality, the whole manifold simply *is*. Whatever asymmetries obtain in such a world require serious explanation. Michael Tooley in fact suggests that causation fundamentally requires a tensed, dynamic, view of time and the world.[83]

[82] Nagel, (1989)
[83] Tooley, (1997), particularly Chapter 4

If we are to accept the possibility of B-series time and the block universe giving the correct description of reality then we must grant that a metaphysics of time and an account of causation have serious consequences for the free will debate. Neither compatibilists nor incompatibilists can afford to follow Van Inwagen's advice above and proceed merely in terms of determinism. Nor can they make the move ubiquitous in the literature and simply refer to determinism as 'causal determinism' – for such talk hides what is at stake: an *account* of asymmetries in determination.

As Hoefer noted, thinking about the metaphysics of time reveals that there is more to incompatibilism than mere determinism. Causal asymmetry does serious work. The incompatibilists have not *missed* this; it is not the case that the concerns of this essay have revealed some lurking lacuna in incompatibilist thinking. Rather, they have been focused upon such an asymmetry all along, but have simply assumed it is obvious and so have failed to argue for it. In light of the fact that the B-series provides a possible description of the nature of time it is necessary that philosophers take the metaphysics of time and causation seriously as they relate to free will. Good arguments against the B-series need to be formulated by those wishing to defend the Consequence Argument; or, failing that, good arguments in favour of causation on the B-series need to be presented. Likewise, compatibilists need to provide good arguments that can explain away apparent asymmetries.

My final point then is to stress an analogy between Kant's analysis of the Copernican revolution and the current state of the free will debate. If the debate has stagnated, perhaps some progress shall be made by examining the assumptions – at times implicit – upon which many of the arguments in the debate turn. The distinction between how the world appears to us and how the world really is has been one of the most profound and difficult problems of philosophy. The pre-Socratic philosophers, Descartes, Kant and the modern phenomenologists have all worried about the division. That such an intractable problem might relate to the seemingly equally vexing problem of free will seems fitting.

5 Conclusion

The problem of free will has proved to be one of the most troubling in the philosophical pantheon of difficult questions. When I first began work on this essay I had hoped that a commitment to a certain metaphysics of time might go some way towards *solving* at least a part of it. I have been somewhat disappointed in that hope. A compatibilism which turns upon the nature of time does not amount to a positive defence of the compatibility of free will and determinism – more is required.

Despite this, I think that concerns about the nature of time as it relates to the free will debate – the fact that determinism does not truly challenge free will unless combined with assumptions about asymmetries in time – are instructive. For such concerns reveal that there is a gap in the Consequence Argument and explain it. Thinking about the debate from the fundamentally tensed perspective of the A-series renders it obvious that such asymmetries exist. As such, incompatibilists have seen no reason to make explicit the assumption that, when combined with determinism, precludes free will. And they have been mystified by compatibilists who abuse a counterfactual dependency sense of agential ability and thereby seem to beg the question. I think that lessons can be learned from Hoefer. Once we adopt what seems to be the correct metaphysics of time the obviousness of the incompatibilists' assumptions disappears. This led Hoefer to think that their assumptions in favour of asymmetry and causation were *mistaken* – and that, merely by a move to the B-series, free will and determinism could be shown to be compatible. This is not the case. In his attempt to defend freedom, Hoefer illustrates the cost of such an attempt. One must give up on causation and face all the attendant strangeness – particularly relating to our actions – that such a move entails. Hoefer, instead of countenancing such strangeness, tried to have his cake and eat it too. He reintroduces causation grounded upon de facto physical asymmetries. In so doing he undermines his thesis: for such a move suggests that perhaps the incompatibilist's assumption in favour of asymmetry was not so mistaken after all. And leaves them room to complete the Consequence Argument and firmly conclude to incompatibilism. If the incompatibilists *cannot* so abuse Hoefer's new asymmetry, he needs to say why they cannot.

I have suggested a different tack. Kant, like Hoefer, sought to argue for the compatibility of free will and determinism based upon a distinction between the world as it appears to us and the world as it is in itself. Kant's ideas in fact seem to have formed the basis for Hoefer's theory. I have suggested that compatibilists, if they make this first Kantian

distinction, should then insist upon it. They should explain away temporally asymmetrical notions such as causation in terms of a kind of idealism. They should make the Kantian-inspired claim that the asymmetries which appear to us do so because of something about *us* – in particular the asymmetric character of deliberation. I remain agnostic as to the success of such an attempt.

However, I believe that the distinction between the A- and B-series of time also shows that much of the debate has been somewhat off the point. For, the understandable A-series assumption of incompatibilists has clouded the fact that at least part of what should be argued about is causation – or at least some serious asymmetry in determination. And more importantly: if what I have said above is correct, then arguments about such asymmetries might point the way to *settling* some of the disagreements between compatibilists and incompatibilists.

My conclusions are therefore rather conservative. In so far as a substantive philosophical position, I claim only that the Consequence Argument and the thesis of incompatibilism are not so certain as they may seem. I argue that there are gaps in the Consequence Argument but that these gaps arise from a very natural perspective on the world and do not, without further arguments, ground any conclusion that the Consequence Argument *fails*. The gaps in the Consequence Argument do not allow us to turn to a defence of robust compatibilism. However, regarding meta-philosophical conclusions about the *debate* between compatibilists and incompatibilists, I suggest that an emphasis should be placed upon investigating the justification of the assumption in the Consequence Argument – an assumption relating to asymmetries in time. I have sketched various positions for both compatibilists and incompatilibists and hope that, if such positions were argued for and analysed, we might go some way to making progress in the free will debate. For, it seems clear that an account of the metaphysics of time plays an important but neglected role in settling the putative challenge to free will from determinism.

Bibliography

Allison, H. (1990), *Kant's Theory of Freedom*, Cambridge University Press.

Anscombe, E. (1981), "Causality and Determination", in *Metaphysics and the Philosophy of Mind,* University of Minnesota Press: pp. 133–147.

Brennan, J. (2007), "Free Will in the Block Universe", *Philosophia* (35) 2: pp. 207-217.

Campbell, J. (2007), "Free Will and the Necessity of the Past", *Analysis,* 67(2): pp. 105-111.

Cartwright, N. (1999), *The Dappled World*, Cambridge University Press.

Dupré, J. (1996), "The Solution to the Problem of the Freedom of the Will", *Noûs,* 30: pp. 385-402.

Dyke, H. (2002), "McTaggart and the Truth About Time", in C. Callender (ed.) *Time, Reality and Experience,* Cambridge University Press.

Earman, J. (1986), *A Primer on Determinism*, D. Reidel Publishing Company.

Earman, J. (2004), "Determinism: What we have Learned and what we Still Don't Know", in J. Campbell, M. O'Rourke, D. Shier (eds.) *Freedom and Determinism*, MIT Press.

Fischer, J. (1986), "Van Inwagen on Free Will", *The Philosophical Quarterly,* 36(143): pp. 252-260.

Fischer, J. (1994), *The Metaphysics of Free Will,* Blackwell Publishers.

Foley, R. (1979), "Compatibilism and Control Over the Past", *Analysis,* 39(2): pp. 70-74.

Forrest, P. (1985), "Backwards Causation in Defence of Free Will", *Mind,* 94(374): pp. 210-217.

Frankfurt, H. (1969), "Alternative Possibilities and Moral Responsibility", *Journal of Philosophy* (66) 23: pp. 829-839.

Gallop, D. (1984), *Parmenides of Elea: Fragments,* University of Toronto Press.

Hoefer, C. (2002), "Freedom from the Inside Out", in C. Callender (ed.) *Time, Reality and Experience,* Cambridge University Press, pp. 201-222.

Hoefer, C. (2004), "Causality and Determinism: Tension, or Outright Conflict?", *Revista de Filosophia*, 29(2): pp. 199-225.

Hoefer, C. (2010), "Causal Determinism", http://plato.stanford.edu/entries/determinism-causal

Honderich, T. (2004), "After Compatibilism and Incompatibilism", in J. Campbell, M. O'Rourke, D. Shier (eds.) *Freedom and Determinism*, MIT Press.

Horgan, T. (1985), "Compatibilism and the Consequence Argument", *Philosophical Studies*, 47(3): pp. 339-356.

Horwich, P. (1987), *Asymmetries in Time*, MIT Press.

Ismael, J. (2007), "Causation, Perspective and Agency", *Psyche*, 13(1).

Kant, I. (2007), *Critique of Pure Reason*, in P. Guyer (translated), A. Wood (eds.), *Immanuel Kant: Critique of Pure Reason*, Cambridge University Press.

Lamb, J. (1977), "On a Proof of Incompatibilism", *The Philosophical Review*, 86(1): pp. 20-35

Lewis, D. (1973), *Counterfactuals*, Blackwell Press.

Lewis, D. (1979), "Counterfactual Dependence and Time's Arrow", *Noûs*, 13: pp. 455-476.

Lewis, D. (1981), "Are we Free to Break the Laws?", *Theoria*, 47: pp. 113-121.

Mackie, J. (1980), *The Cement of the Universe: A Study of Causation*, Oxford University Press.

Mackie, P. (2003), "Fatalism, Incompatibilism and the Power to do Otherwise", *Noûs*, 37(4): pp. 672-689.

McKenna, M. (2009), "Compatibilism" http://plato.stanford.edu/entries/compatibilism

McTaggart, J. (1908), "The Unreality of Time", *Mind* (17) 68: pp. 457-474.

Mellor, D. (1985), *Real Time*, Cambridge University Press.

Nagel, T. (1989), *The View from Nowhere*, Oxford University Press.

Pereboom, D. (2001), *Living Without Free Will*, Cambridge University Press.

Popper, K. (1956), "The Arrow of Time", *Nature*, 177: pp. 538.

Price, H. (1996), *Time's Arrow and Archimedes' Point: New Directions for the Physics of Time*, Oxford University Press.

Russel, B. (1912), "On the Notion of Cause", *Proceedings of the Aristotelian Society*, 13: pp. 1-26.

Sartorio, C. (2014), "The Problem of Determinism and Free Will is Not the Problem of Determinism and Free Will", in A. Mele (ed.) *Surrounding Free Will: Philosophy, Psychology, Neuroscience*, Oxford University Press.

Savitt, S. (2013), *Being and Becoming in Modern Physics*, http://plato.stanford.edu/entries/spacetime-bebecome/

Sider, T. (2001), *Four-Dimensionalism*, Oxford University Press.

Tooley, M. (1997), *Time, Tense and Causation*, Oxford University Press.

Van Inwagen, P. (1980), "Compatibilism and the Burden of Proof", *Analysis*, 40(2): pp. 98-100.

Van Inwagen, P. (1983), *An Essay on Free Will*, Oxford University Press.

Van Inwagen, P. (1990), "Logic and the Free Will Problem", *Social Theory and Practice*, 16(3): pp. 277-290.

Van Inwagen, P. (2004), "Freedom to Break the Laws", *Midwest Studies in Philosophy*, 28: pp. 334-350.

Van Inwagen, P. (2008), "How to Think about the Problem of Free Will", *Journal of Ethics* (12) 3: pp. 327-341.

Van Inwagen, P. (2008), *Metaphysics*, Oxford University Press.

Vihvelin, K. (1991), "Freedom, Causation, and Counterfactuals," *Philosophical Studies*, 64: pp. 161–184.

Vihvelin, K. (2011), "Arguments for Incompatibilism", http://plato.stanford.edu/entries/incompatibilism-arguments/