The Logic of Past-Alteration

Abstract: Is it possible to change the past – to make something that *has* happened *not* have happened? Past-alteration is widely believed to be 'logically impossible'. But despite this, there have been few attempts to actually apply logical resources to the question of whether it is possible to change the past. This paper articulates a novel tense logic and uses it to argue that past-alteration is possible – with just a *single* dimension of time – so long as it's possible for time to have a certain kind of structure: one which is 'backwards-branching' and in which the relation of precedence 'comes apart' from the relation of succession. The remainder of the paper explores some of the metaphysical implications of this account of what it is to change the past.

1. Introduction

There are many reasons to want to travel to the past – see the dinosaurs, talk to a prophet, find out what the Voynich manuscript was really all about. But ask most people what they'd do with a time machine and they'll tell you not just of their hopes of *seeing* the past, but also of *changing* it – of killing baby Hitler to prevent the Holocaust, warning previous generations of impending environmental disasters, maybe thinking twice before sending that passive-aggressive email to Kevin in marketing yesterday. Are such hopes really coherent? Is it possible to change the past?

Many philosophers believe that the answer to this question is *no*. After all, they point out, these events (Hitler's birth, the email to Kevin, and so on) have, by hypothesis, *already happened*; the idea that there is anything we could do now to make these things *not* have happened is "logically incoherent" (Fulmer 1983: 33), "senseless" (Brown 1985: 85), "radically absurd" (Flew 1954: 57), "a logical impossibility" (Hospers 1967: 177). The past is unshakeable; it cannot be altered.

Against this prevailing sentiment, some philosophers have recently suggested that we can make sense of past-alteration if we are willing to postulate an extra dimension of time.[1] But for all the talk of past-alteration's 'logical impossibility', there have been few attempts to

[1] See, e.g., Meiland (1974), Goddu (2003, 2011), van Inwagen (2010), Hudson and Wasserman (2010), Bernstein (2017).

actually apply logical resources to the question of whether past-alteration is possible.[2] This is what I propose to do in this paper. I start by showing that past-alteration scenarios violate two important tense-logical principles: (LIN-P) (roughly, the claim that *whatever will have been the case is sometime the case*) and (GP) (roughly, the claim that *whatever is the case will always have been the case*). I'll then construct a tense logic which allows for violations of (LIN-P) and (GP), and use it to argue that past-alteration is possible – with just a *single* dimension of time – if, but only if, it's possible for time to be 'backwards-branching' in structure and for the relation of precedence to 'come apart' from the relation of succession (in a sense that will be made more precise shortly). The rest of the paper will explore some of the metaphysical implications of this novel account of what it is to change the past.

2. Clarifying the Target

Before we begin, we should distinguish three quite different things that might be meant by 'changing the past'. On the first interpretation, to change the past is to have a *causal effect* on the past; to cause something which did in fact happen to have happened. If JFK was killed, not by Oswald, but by a time-traveller from the future, then the time-traveller 'changes the past' in this sense – the past is the way it is in part because of the time-traveller's future decision to travel back in time. I accept, along with many others,[3] that it's possible to change the past in this sense. But this is not the kind of past-alteration I am interested in. The kind of past-alteration I am interested in involves more than simply causing the past to be the way it is; it involves making the past *different* from how it (now) is.

As Vranas (2005) points out, though, even this definition of past-alteration is ambiguous between two different readings. On the first reading, changing the past involves making the past different to how it *actually* is. I also accept – again, along with many others[4] – that it's possible to change the past in this sense. Suppose that a time-traveller from the future appeared in her time machine in late 19th-Century Austria, with the sole intention of killing Hitler before his rise to power. There's a sense in which she *could* have killed Hitler – she had the motive and the means, her weapon was loaded, the line of sight was good, and so on. Of

[2] A notable exception is Wasserman (2017: 102-6), who in a suggestive passage correctly identifies that any logic of past-alteration would have to permit violations of the (GP) axiom of standard tense logic. Wasserman doesn't actually provide such a logic, however. Moreover, he appears to presuppose that the option of constructing a tense logic that permits violations of (GP) to make sense of past-alteration is one which is only available to A-theorists, though as I argue below (§5) the strategy is in fact perfectly neutral with respect to the debate between A-theorists and B-theorists. Nevertheless, this paper can be seen as a development of ideas already nascent in Wasserman's discussion.

[3] See, e.g., Brier (1973: 363), Dwyer (1977) and Horwich (1975: 435-7). For a discussion of common objections to the claim that time-travellers could causally affect the past, see Wasserman (2017: ch.5).

[4] See especially Lewis (1976: 149-50).

course we know that she *didn't*, in fact, kill Hitler, since Hitler actually died in 1945 by his own hand. But there is still a possible world, relevantly similar to the actual world, in which this time-traveller kills Hitler, thereby making the past different from how it actually is.[5] Again, though, this is not the kind of past-alteration I am interested in. What I am interested in is not whether there is a possible world in which someone makes the past different from how it actually is; what I am interested in is whether there is a possible world in which someone makes the past different from how it used to be *in that world*.

To illustrate, consider the following case:

> *Back to the Future*
> *It's 1.22am on the 26th October 1985. Local eccentric Doc Brown has persuaded his friend Marty to help him test out his latest invention – a time-machine made from a DeLorean car. The necessary 1.21 gigawatts of energy is supplied by a plutonium nuclear reactor, the fuel for which Doc has stolen from a group of Libyan nationalists. Unfortunately the Libyans have tracked Doc down and, seconds earlier, they shot him dead. Marty escaped into the DeLorean and is now accelerating away from the scene. Very soon, he will hit 88mph, the flux capacitor will activate, and Marty will travel back to the last time entered into the computer: 06.38 on the 5th November 1955. After that happens, the past will be very different to how it is now – Marty will have played Johnny B. Goode at a high school prom, met his mother as a teenager, and warned Doc about the terrorists, thereby saving his life.*[6]

Notice that, in *Back to the Future*, the past is not the way it *is* because of anything that Marty will do in the future. Nor is *Back to the Future* merely a situation in which Marty could have acted differently in the past to how he actually acted. Instead what *Back to the Future* represents, or purports to represent, is a situation in which the past (including the distant past[7]) will soon be *different*, because of what Marty is about to do, to how it *is* now. The question I am interested in is whether such situations are metaphysically possible.

---

[5] According to Vihvelin (1996), it's not sufficient for one to be *able* to φ, in the ordinary sense, that there is some relevantly similar possible world in which one φs; it must also be the case that *were one to try* to φ, one would (or at least might) succeed. Vihvelin argues on these grounds that although time-travellers might be able to, say, kill baby Hitler, they wouldn't be able to kill their grandfathers, or their past selves, because (roughly speaking) the worlds where they try to kill their past selves and succeed are all further away from actuality than the worlds where they try to kill their past selves and fail; c.f. Sider (2002) and Vranas (2010) on this point.

[6] Note that *Back to the Future* is set at a particular time. Here and throughout, when I say that a sentence is true in a case like *Back to the Future*, I will mean that it is true *at the time at which the case is set*, in this case, 1.22am on 26th October 1985.

[7] Of course, there's a trivial sense in which every action makes the past different to how it was – the past didn't used to contain that action, and now it does. Clearly, though, *Back to the Future* involves more than just this trivial kind of past-alteration.

As I noted at the outset, the received wisdom among philosophers is that cases like *Back to the Future* are not metaphysically possible. Even the very first explicit written discussion of past-alteration, in Enrique Gaspar's *El Anacronópete*, makes clear that although "[w]e may be present to witness facts consummated in preceding centuries…we may never undo their existence" (Gaspar 1887: 59).[8] Indeed, time-travel fiction is replete with cautionary tales of time travellers who, in trying to prevent past events from happening, end up causing those very events to occur.[9] The message to those foolish enough to think they can change the past seems clear: don't bother. At *best* you'll fail; and at worst, you'll end up causing the very events you're trying to prevent!

This attitude strikes me as premature, however. *Back to the Future* is not *obviously* contradictory; it doesn't say of any fact that it both does and doesn't obtain, for example. Before we can determine whether the situation it describes is metaphysically possible, then, we need to know more about what would have to be true about the temporal structure of the universe for such a situation to obtain. This is my goal in what follows. My strategy, outlined in the next section, will be to use the resources of tense logic to characterise, in precise terms, the metaphysical principles I take to be at stake in the debate over whether past-alteration is possible. I start with a brief introduction to tense logic, before applying the framework to cases like *Back to the Future*.

3. Past-Alteration in Tense Logic

Standard tense logic (TL), originally developed over a series of papers in the 1950s and 60s by Arthur Prior,[10] extends classical propositional logic with two operators, **P** and **F**, intuitively paraphrased as 'It has at some time been the case that' and 'It will at some time be the case that', respectively. Other operators can then be defined in terms of **P** and **F**, including:

$\mathbf{H}\varphi =_{df} \neg\mathbf{P}\neg\varphi$ ("It has always been the case that")
$\mathbf{G}\varphi =_{df} \neg\mathbf{F}\neg\varphi$ ("It will always be the case that")
$\mathbf{S}\varphi =_{df} (\varphi \vee \mathbf{P}\varphi \vee \mathbf{F}\varphi)$ ("It's sometime the case that")
$\mathbf{A}\varphi =_{df} (\varphi \wedge \mathbf{H}\varphi \wedge \mathbf{G}\varphi)$ ("It's always the case that")

The semantics of TL are modelled on the Kripke semantics for modal logic. Let a *temporal frame* be an ordered pair $\langle T, \prec \rangle$, where T is a non-empty set and $\prec$ is a relation over T (that is,

[8] Thanks to Michael Main and his 'Big List of Time Travel Adventures', available at https://www.storypilot.com/, for this historical detail.

[9] See, e.g., *One Life, Furnished in Early Poverty* from the Twilight Zone series.

[10] Prior (1957, 1967, 1969). My presentation is based on Goranko and Galton (2015).

a set of ordered pairs of elements of T). Informally, one can think of T as the set of 'times' and $\prec$ as the relation of 'temporal precedence', though these informal characterisations should not be understood as constraints on what counts as a temporal frame. A *temporal model* is an ordered triple $M = \langle T, \prec, V \rangle$, where V is a function that assigns a subset of T to every atomic formula of TL (informally, one can think of $V \in M$ as the function which assigns to every atomic formula the set of times at which, according to M, that formula is true). The truth of a formula φ in a model M at a time t is defined as follows (where '$M, t \vDash \varphi$' abbreviates 'φ is true in M at t'):

- If φ is an atomic formula of TL, $M, t \vDash \varphi$ iff $t \in V(\varphi)$.
- $M, t \vDash \mathbf{P}\varphi$ iff $M, t` \vDash \varphi$ for some $t`$ such that $t` \prec t$.
- $M, t \vDash \mathbf{F}\varphi$ iff $M, t` \vDash \varphi$ for some $t`$ such that $t \prec t`$.

To the above we should also add, of course, the usual semantic clauses for the propositional connectives. A formula φ is a *theorem* of TL iff $M, t \vDash \varphi$ for all M and t.

It's important to emphasise that TL is, in-and-of-itself, metaphysically neutral. Merely by using the resources of tense logic to regiment one's claims about what was and will be the case, one doesn't commit oneself to any particular metaphysics of time, just as by using the resources of modal logic to regiment one's claims about what is necessary or possible one doesn't commit oneself to any particular metaphysics of modality. Nevertheless, TL represents a powerful formal framework within which competing metaphysical hypotheses can be precisely stated and their relative merits systematically assessed.

To see this, let's go back to *Back to the Future*. Two two distinctive features of *Back to the Future* are worth highlighting. First, there are things which *will have* been true in *Back to the Future* which are *never* true now. For example, the following is true in *Back to the Future*:

> (1) Marty has never played and will never play Johnny B. Goode at a high-school prom, but it will be the case that he did.

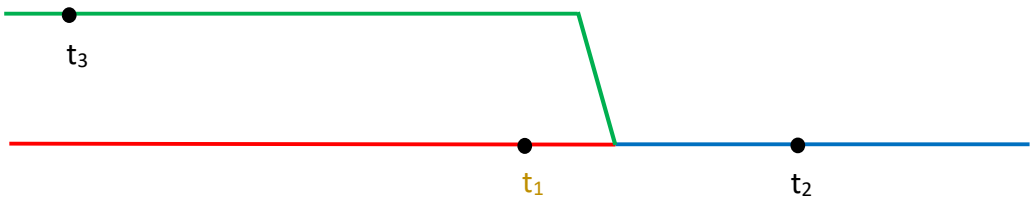
Here's the same claim formalised in TL (where 'p' stands for the sentence 'Marty is playing Johnny B. Goode at a high-school prom'):

> (1*) $\neg\mathbf{S}p \wedge \mathbf{FP}p$

(1*) is a violation of the following schema:

> (LIN-P): $\mathbf{FP}\varphi \rightarrow \mathbf{S}\varphi$ ("Whatever will have been the case is sometime the case")

In TL, (LIN-P) corresponds to the claim that $\prec$ is *linear in the past*, i.e. that for all $t, t`, t^* \in T$, if $t \prec t^*$ and $t` \prec t^*$, then either $t \prec t`$ or $t` \prec t$ or $t = t`$ (or in other words, if two distinct times have a successor in common, one must precede the other). This means that any model of TL in which (1) comes out true would have to be 'backwards-branching' in structure. Consider the following model, for example:



M1

Each point in this diagram represents a time, and the order of precedence runs from left to right (so for example $t_1 \prec t_2$ and $t_3 \prec t_2$, but $t_3 \not\prec t_1$, $t_1 \not\prec t_3$, $t_3 \neq t_1$). t_1 represents the time at which *Back to the Future* is set (1.22am on the 26th October 1985), t_2 a time after Marty hits 88mph, and t_3 a time in the 'altered' past. Suppose that p (the sentence 'Marty is playing Johnny B. Goode at a high-school prom') is true at t_3 but not true at any time in the red or blue branches. Then it's easy to show that (1) is true at t_1: $\neg \mathbf{S}p$ is true at t_1, since there is no time t such that either $t \prec t_1$ or $t_1 \prec t$ or $t = t_1$ at which p is true (i.e. p is false everywhere along the red and blue branches); and $\mathbf{FP}p$ is true at t_1, since p is true at t_3, and there is a time – e.g. t_2 – such that $t_1 \prec t_2$ and $t_3 \prec t_2$. More informally, (1) turns out true at t_1 because although there is no time in t_1's future or past at which Marty is playing Johnny B. Goode at a high-school prom, there is a time in t_1's future which includes as part of *its* past a time at which Marty is playing Johnny B. Goode at a high-school prom.

Unfortunately, though, we don't yet have a model on which *Back to the Future* is true. This is because there are also things which *are* true in *Back to the Future* which *will always* have been false. For example, the following is true in *Back to the Future*:

(2) Doc Brown's body is ridden with bullets, but it will be the case that Doc's body has never been ridden with bullets.

Formalised in TL (where 'q' stands for 'Doc Brown's body is ridden with bullets'):

(2*) $q \wedge \mathbf{FH}\neg q$

(2*) is a violation of the following schema:

(GP): $\varphi \rightarrow \mathbf{GP}\varphi$ ("Whatever is the case will always have been the case")

And the problem is that (GP) is simply a *theorem* of TL – every one of its instances is true at every time in every model, regardless of the formal properties of $\prec$.[11] It follows, therefore, that (2) is a TL-contradiction, and hence that *Back to the Future* is a TL-contradiction too.

Some people might be tempted to point to this as conclusive evidence that past-alteration is impossible. But this would be too hasty. TL, remember, is merely a formal tool – it is not in-and-of-itself an arbiter of what is necessary or possible. As Wasserman notes, standard tense logic arguably "begins with the assumption that the past cannot change", and so "cannot provide an *independent* reason to give up that possibility" (Wasserman 2017: 104). To quote Prior himself:

> The logician must be rather like a lawyer…in the sense that he is there to give the metaphysician, perhaps even the physicist, the tense-logic that he wants, provided that it be consistent. He must tell his client what the consequences of a given choice will be…and what alternatives are open to him; but I doubt whether he can, *qua* logician, do more. We must develop, in fact, alternative tense-logics, rather like alternative geometries; though this is not to deny that the question of what sort of time we actually live in, like the question of what sort of space we actually live in, is a real one, or that the logician's exploration of the alternatives can help one to decide it. (Prior 1967: 59)

I propose to follow Prior's advice. If we're to have any chance of understanding what a world in which past-alteration occurs would actually be like, we need a tense logic that is weaker than TL, one on which (GP) is not automatically a theorem. In fact there's a fairly straightforward way of achieving this. The reason why (GP) turns out to be a theorem of TL is that the semantics of **P** and **F** in TL are both defined in terms of the same relation, $\prec$. What we need instead is *two* relations: a relation of temporal *precedence* and a relation of temporal *succession*. Let $\mathrm{TL}^{(\text{-})}$ be a tense logic whose models are ordered quadruples, $\langle \mathrm{T}, \prec, \succ, \mathrm{V} \rangle$, where T is a set, $\prec$ and $\succ$ are relations over T, and V is a function that assigns a subset of T to every atomic formula of $\mathrm{TL}^{(\text{-})}$. The truth of a formula φ of $\mathrm{TL}^{(\text{-})}$ in model M is then defined as follows:

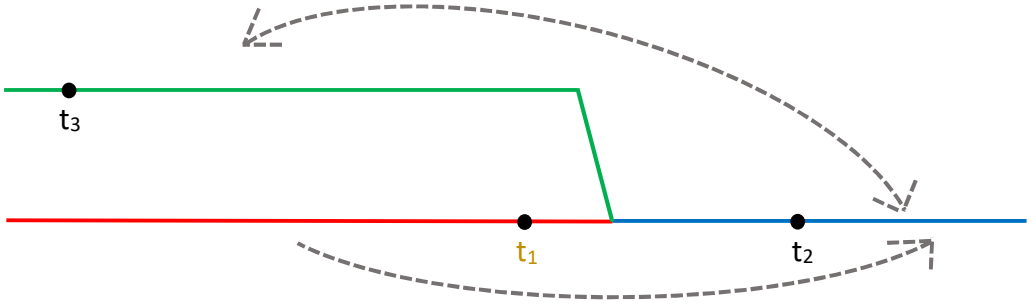
- If φ is an atomic formula of $\mathrm{TL}^{(\text{-})}$, $\mathrm{M}, t \vDash \varphi$ iff $t \in \mathrm{V}(\varphi)$.
- $\mathrm{M}, t \vDash \mathbf{P}\varphi$ iff $\mathrm{M}, t` \vDash \varphi$ for some $t`$ such that $t` \prec t$.

[11] *Proof*: Suppose for *reductio* that there is some φ, t and M such that $\mathrm{M}, t \vDash \varphi$ but $\mathrm{M}, t \nvDash \mathbf{GP}\varphi$. Then, by the semantic clause for **F**, there is some $t`$ such that $t \prec t`$ and $\mathrm{M}, t` \nvDash \mathbf{P}\varphi$; and so, by the semantic clause for **P**, there is no $t^*$ such that $t^* \prec t`$ and $\mathrm{M}, t^* \vDash \varphi$. But there is such a t^* – namely, t! Contradiction. So there is no φ, t and M such that $\mathrm{M}, t \vDash \varphi$ but $\mathrm{M}, t \nvDash \mathbf{GP}\varphi$. Hence $\mathrm{M}, t \vDash \varphi \rightarrow \mathbf{GP}\varphi$, for all φ, t and M.

- M, $t \vDash \mathbf{F}\varphi$ iff M, $t` \vDash \varphi$ for some $t`$ such that $t` \succ t$.

Notice that while the first two clauses are the same as those for TL, the third clause for **F** is defined in terms of $\succ$, not, as it is in TL, in terms of $\prec$.[12]

In TL$^{(-)}$, (GP) corresponds to the claim that precedence is the *inverse* of succession, i.e. that for all t and $t`$, if $t` \succ t$, then $t \prec t`$. This fits with our ordinary understanding of precedence and succession, of course – usually, a time precedes every time which succeeds it. Crucially, however, TL$^{(-)}$ allows for models in which this is not the case. Consider the following model, for example:



M₂

The dotted lines here indicate relations of precedence and succession – so, all the times in the green branch precede and are succeeded by all the times in the blue branch (and in particular, $t_3 \prec t_2$ and $t_2 \succ t_3$), but while every time in the blue branch succeeds every time in the red branch, no time in the red branch precedes any time in the blue branch (and in particular, $t_2 \succ t_1$, but $t_1 \not\prec t_2$). It's easy to show then that both (1) and (2) are true in $\mathbf{M_2}$ at t_1. (1) is true in $\mathbf{M_2}$ for much the same reason it was in $\mathbf{M_1}$; namely, that p is false at all times t such that $t \prec t_1$, $t \succ t_1$ or $t = t_1$, but there *is* a time $t \succ t_1$ and a time $t` \prec t$ such that $p$ is true at $t`$ (in particular, p is true at t_3, and $t_3 \prec t_2 \succ t_1$). (2), meanwhile, is true in $\mathbf{M_2}$ at t_1 because q is true at t_1, and there is a time $t \succ t_1$ such that q is false at all times $t`$ such that $t` \prec t$ (in particular, $t_2 \succ t_1$, and q is false at all times preceding t_2). More informally, (2) turns out true in $\mathbf{M_2}$ at t_1 because there is a time in t_1's future which *doesn't* include t_1, or any other time at which Doc Brown's body is ridden with bullets, as part of its past.

[12] Although $\prec$ and $\succ$ are distinct accessibility relations, our preferred logic can still impose constraints on how they are able to interact. For example, although the minimal logic of TL$^{(-)}$ allows for models where $t \prec t`$ and $t \succ t`$, as well as models with larger 'loops' (e.g. models where $t_1 \prec t_2$, $t_3 \succ t_2$, and yet $t_3 \prec t_1$), there might be good reasons to add additional axioms to our logic that specifically rule out such models (which would still allow for violations of (GP), i.e. models on which $t \prec t`$ but not $t` \succ t$). Thanks to [REDACTED] here.

Let me summarise where we've got to so far. Past-alteration scenarios like *Back to the Future* can be characterised by their failure to conform to two important principles – (LIN-P), the principle that *whatever will have been the case is sometime the case*, and (GP), the principle that *whatever is the case will always have been the case*. Although there are models of TL which violate (LIN-P), (GP) is simply a theorem of standard tense logic. But there is a weaker logic we can construct, $TL^{(-)}$, on which (GP) is not a theorem. In this logic, cases like *Back to the Future* are not contradictory – there are models, like $\mathbf{M_2}$, on which they come out true, ones which are 'backwards-branching' in structure (i.e. where two distinct times can have a successor in common without either preceding or succeeding the other), and on which precedence 'comes apart' from succession (i.e. where one time can succeed another without the second preceding the first). I conclude, therefore, that past-alteration is metaphysically possible if, but only if, it is metaphysically possible for time to be backwards-branching and for succession to come apart from precedence, in the senses described above. This is, as we'll see, an entirely novel characterisation of what is at stake in the debate over whether past-alteration is possible. Yet it falls out naturally from formalisations of past-alteration scenarios in tense logic.

Perhaps it just seems obvious to you that succession couldn't come apart from precedence in this way. *Surely*, you might be thinking, if one time succeeds another, the second must precede the first, just as if one person is taller than another the second must be shorter than the first. But it's important to remember that 'precedence' and 'succession' in this context are technical terms, mere labels for the accessibility relations governing the semantics of the tense operators – we cannot simply rely on our intuitions about how they ought to interact. The real substantive question here is whether violations of (GP) and (LIN-P) are metaphysically possible. And this, it seems to me, is a genuinely open metaphysical question, to be resolved in the usual ways such questions are resolved – by teasing out the consequences of different answers and examining how they interact with our background metaphysical commitments. I will do some of this work in later sections. But first, it will be worth examining how the approach developed above applies to the equally interesting – though comparatively neglected[13] – phenomenon of *future*-alteration.

4. Future-Alteration in Tense Logic

It's often alleged that unlike travel to the past, "[t]ravel to the future raises no conceptual problems" (van Inwagen 2010: 3). This is a mistake, however; in fact, exactly the same considerations apply to travel to the future as do to travel to the past. Presumably no-one

[13] An exception is Hudson and Wasserman (2010); see also Wasserman (2017).

would deny that a time-traveller could *cause* something which will in fact happen far in the future to happen; the more interesting question is whether anyone could *alter* the future, i.e. prevent something which *will* in fact happen from happening (or cause something which *won't* happen to happen). Consider the following story, for example:

Back to the Future II
It's the 21st October 2015. Doc Brown and Marty have recently arrived in the DeLorean from the 26th October 1985. Under Doc's instructions, Marty has donned an iridescent baseball cap to disguise himself as his son, Marty Jr. Soon Marty will head to Café 80s (one of those "not-well-executed nostalgic-themed restaurants") to prevent Griff (grandson of Marty's father's tormenter Biff) from meeting Marty Jr.; Marty Jr. will then go on to live a full and happy life. Before 1985, however, the future was very different to how it is now – back then, it was going to be the case that, in 2015, Marty Jr. will be coaxed by Griff into committing robbery, caught by the police, and convicted to 15 years in jail.

As with *Back to the Future*, two distinctive features of *Back to the Future II* are worth highlighting for our purposes. First, there are things which *were going to be* the case in *Back to the Future II* which are *never* true now. For example, the following seems true in *Back to the Future II*:

(3) Marty Jr. was going to be convicted of robbery, though he has never and will never be convicted of robbery.

Here is (3) formalised in tense logic (where '*r*' stands for the sentence 'Marty Jr. is convicted of robbery'):

(3*) $\mathbf{PF}r \wedge \neg\mathbf{S}r$

(3*) is a violation of the following schema:

(LIN-F): $\mathbf{PF}\varphi \rightarrow \mathbf{S}\varphi$ ("Whatever was going to be the case is sometime the case")

In TL, (LIN-F) corresponds to the claim that $\prec$ is *linear in the future*, i.e. that for all $t, t`, t^* \in$ T, if $t^* \prec t$ and $t^* \prec t`$, then either $t \prec t`$ or $t` \prec t$ or $t = t`$ (or in other words, if two distinct times have a predecessor in common, one must precede the other). This means that any model of TL in which (3) comes out true would have to be *forwards*-branching in structure.

As with past-alteration, however, there are also things which *are* true in *Back to the Future II* which *were never going to be* true before 1985. For example, the following is true in *Back to the Future II*:

(4) Marty is wearing an iridescent cap, though it used to be the case that he will never wear an iridescent cap.

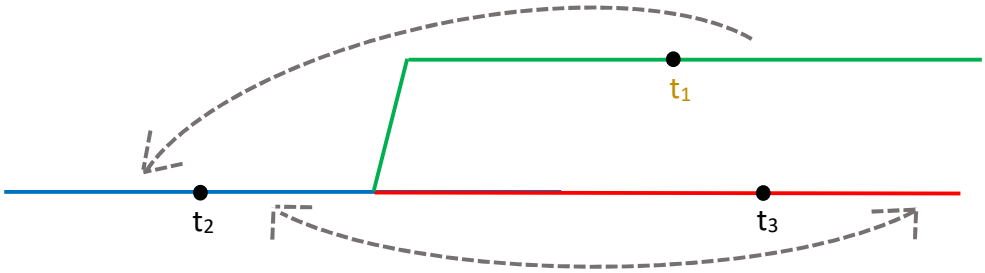
Formalised in tense logic (with '*s*' standing for 'Marty is wearing an iridescent cap'):

(4*) $s \wedge \mathbf{PG}\neg s$

(4*) is a violation of the following schema:

(HF): $\varphi \rightarrow \mathbf{HF}\varphi$ ("Whatever is the case was always going to be the case"

And (HF) is a theorem of TL – all its instances are true in every model at every time, regardless of the formal properties of $\prec$. It follows that (4), and by extension *Back to the Future II*, are TL-contradictions; there are no TL-models on which they're true. But there *is* a TL$^{(-)}$-model on which (4) is true. Consider the following model, for example:



M3

As before, the dotted lines indicate relations of precedence and succession, so all the times in the blue branch precede and are succeeded by all the times in the red branch (and in particular, $t_2 \prec t_3$ and $t_3 \succ t_2$), but while every time in the blue branch precedes every time in the green branch, no time in the green branch succeeds any time in the blue branch (in particular, $t_2 \prec t_1$, but $t_1 \not\succ t_2$). Suppose *r* is true at t_3 but not at any time in the blue or green branches, and suppose *s* is true at t_1 but not at any time in the blue or red branches. It's easy then to show that both (3) and (4) are true in $\mathbf{M_3}$ at t_1. (3) is true at t_1 because there is a time $t \prec t_1$ and a time $t` \succ t$ such that *r* is true at $t`$, but *r* is false at all times *t* such that $t \prec t_1$, $t \succ t_1$ or $t = t_1$ (in other words, although there is no time in t_1's future or past at which Marty Jr. is convicted of robbery, there is a time in t_1's past which includes as part of *its* future a time at which Marty Jr. is convicted of robbery). (4), meanwhile, is true at t_1 because *s* is true at t_1, and there is a time $t \prec t_1$ such that *s* is false at all times $t`$ such that $t` \succ t$ (in other words, although Marty is wearing an iridescent cap at t_1, there is a time in t_1's past which doesn't include t_1, or any other time at which Marty is wearing an iridescent cap, as part of its future). We can conclude, therefore, that future-alteration is metaphysically possible *only if* it is metaphysically possible

for time to be 'forward-branching' in structure (i.e. for two distinct times to have a predecessor in common without either preceding or succeeding the other) and for precedence to 'come apart' from succession (i.e. for one time to precede a second without the second succeeding the first).

For the most part, I will continue to focus on past-alteration rather than future-alteration in what follows. As we've seen in this section, however, there are important parallels between travelling to the past and travelling to the future, and much of what I will have to say about past-alteration will apply, *mutatis mutandis*, to future-alteration too.

5. Interpreting the Models

So far, I have presented tense logical models on which past-alteration scenarios, like *Back to the Future* (and future-alteration scenarios, like *Back to the Future II*), come out true. But how should we, *qua* metaphysicians, interpret these models? What would a world accurately represented by such models actually be like?

The answer to this question will depend, to some extent, on one's background metaphysics of time. Let's start by thinking about how a *B-theorist* would interpret models like $\mathbf{M_2}$. According to the B-theory, all times are metaphysically on a par. In particular, there is no objectively present moment – the best we can say is that every time is present relative to itself. Reality as a whole is a vast, four-dimensional manifold, with concrete objects spread throughout that manifold. Speaking unrestrictedly, the B-theorist insists, there are dinosaurs and dodos and (presumably) female US presidents – it's just that most of our utterances have their quantifiers restricted to the set of things located at the time the utterance is made. There's a sense, then, in which the state of reality *as a whole* is fixed for all time, on the B-theory;[14] to say that things 'change over time' is really just to say that there's a qualitative difference between different parts of reality along the temporal dimension, "just as a 'change' in scenery from east to west

[14] There's an interesting debate I'm sidestepping here about how the B-theorist should think about the interaction between tense operators and quantifiers. On what might be called the 'standard' view (see, e.g., Sider (2001, 2006)), tense operators restrict the domains of quantifiers within their scope to the set of things located at the relevant time – for example, '**P**(There are dodos)' is true at *t* iff there is a time preceding *t* at which dodos are located. But as Deasy (forthcoming) convincingly argues, there are serious problems with this view. Instead, Deasy suggests, B-theorists should think of the tense operators as *logically redundant* when the sentences in their scope are purely qualitative – for example, '**P**(There are dodos)' is logically equivalent to 'There are dodos', on the B-theory (so long as they are both evaluated in the same context). This debate needn't concern us, however, since the tensed sentences at issue for us (namely, (1)-(4)) are *not* purely qualitative. Even if it follows from the B-theory that 'There are dodos' is true at every time if it is true at some time, it doesn't follow from the B-theory that 'Marty is playing Johnny B. Goode' is true at every time if it is true at some time (though B-theorists will disagree, of course, about whether such sentences should be analysed in terms of temporal parts, temporal counterparts, in some other way, or not at all).

is a qualitative difference between the eastern and western spatial parts of the landscape" (Lewis 1976: 145).

Suppose the B-theorist is correct. Then what models like $\mathbf{M_2}$ represent is a four-dimensional concrete universe, albeit one with a non-standard spatiotemporal structure. In 'standard' universes, for example, every time-slice either precedes or succeeds every other. But in the universe represented by $\mathbf{M_2}$, the time-slice at t_1 neither precedes nor succeeds the time-slice at t_3. This means that although it's true at t_1 that an event of Marty playing Johnny B. Goode at a high-school prom *exists* (speaking unrestrictedly, in some bit of the universe), it's also true at t_1 that this event never has occurred and will never occur, since the bit of the universe in which Marty is playing Johnny B. Goode at a high-school prom is temporally inaccessible from the bit of the universe located at t_1.

Some might worry that $\mathbf{M_2}$ on the B-theoretic interpretation seems more like a case of *swapping one past for another* than it does a case of genuine past-*alteration*. "Marty doesn't actually *change* the red branch," it might be objected; "nor, indeed, does he change the green branch. He just moves from a bit of reality where the red branch is in his past to a bit of reality where the green branch is in his past."[15] It's important to remember, though, that change for the B-theorist *just is* variation along the temporal dimension. For the past to change between t_1 and t_2, for the B-theorist, is just for there to be a difference between the parts the universe which are precedence-related to t_1 and those which are precedence-related to t_2; and this is exactly what the B-theorist thinks is represented by $\mathbf{M_2}$. You might well continue to feel like something fishy is going on here – that notwithstanding the fact that there is a difference between which bits of the universe are precedence-related to t_1 and which bits are precedence-related to t_2, there's still an important sense in which the past doesn't *really* change on the B-theoretic interpretation of $\mathbf{M_2}$. But if so, your complaint is with the B-theorist's conception of change *simpliciter*, not with their conception of past-alteration *per se*.

Consider next, therefore, how models like $\mathbf{M_2}$ would be interpreted by an *A-theorist* about time. The A-theorist objects to the B-theory on the grounds that it cannot account for *genuine* change, since how things are, unrestrictedly speaking, is fixed for all time on the B-theory. When an A-theorist considers the vast, four-dimensional concrete reality postulated by the B-theorist, they only see more of what there *is*, not what there was or will be. According to the A-theory, tensed claims are true when uttered at some times and false when uttered at others, not because those utterances express different things about different bits of an unchanging concrete reality, but because how things are – unrestrictedly speaking, with our quantifiers

[15] See Smith (1997: 365-6) for a version of this objection.

ranging over everything – *changes over time*. What's interesting about past-alteration scenarios like *Back to the Future* is that they involve genuine change, not just in how things *are*, but also in how things *were*. We can model such scenarios by means of mathematical objects like $\mathbf{M_2}$ with 'branching' structures; but as the A-theorist thinks of them, these models shouldn't be interpreted as carrying any commitment to the concrete existence of the 'branches' themselves. All such models imply is that there are times at which things will have been different from how they were – and it's not the case, according to the A-theory at least, that every way concrete reality *used to* or *will* be is a way a part of concrete reality, unrestrictedly speaking, *is*.

It's important to note, however, that the A-theory is compatible with many different views about the persistence of objects over time. Some A-theorists, of course, think that only *present* things exist. According to them, for example, although Socrates used to exist, he exists no longer. But the A-theory is also compatible with the following view:

Permanentism: $\mathbf{A}\forall x\mathbf{A}\exists y\ x{=}y$ ("Always, everything is always something")

According to Permanentism, Socrates still exists, given that he used to exist – there is still something that is Socrates. But this thing isn't a person, or a philosopher, or even concrete, though of course it *used* to be all these things. According to the A-theory, there are no dinosaurs, even speaking unrestrictedly – there is no bit of the universe in which dinosaurs (still) exist. But there *used* to be dinosaurs, of course – the universe *used* to be such that it contained dinosaurs. If Permanentism is true, therefore, it follows that there are *things which used to be* dinosaurs – 'former-dinosaurs', as we might call them. Mere former-dinosaurs aren't dinosaurs, any more than mere alleged crimes are crimes. Indeed former-dinosaurs aren't even concrete, insofar as they are not (any longer) located anywhere in space. But they still exist, according to Permanentism – there is still something they are identical to.

Interestingly, however, even Permanentism leaves open some important questions about past-alteration cases like *Back to the Future*. Consider the thing which, at t_3, is the event of Marty's playing Johnny B. Goode – call this thing *Guitar Solo*. Consider also the thing which, at a time shortly before t_1, is the event of Doc Brown being shot by terrorists – call this thing *Terrorist Attack*. Does *Guitar Solo* exist at t_1, i.e. even before the past is altered? And does *Terrorist Attack* still exist at t_2, i.e. even after the past has been altered? It doesn't follow from Permanentism either that *Guitar Solo* exists at t_1 or that *Terrorist Attack* exists at t_2, since at t_1, *Guitar Solo* never has occurred and never will occur, and similarly for *Terrorist Attack* at t_2.

We can therefore distinguish two different views about past-alteration, both of which are consistent with the conjunction of Permanentism and the A-theory. According to the first view, something exists at every time if and only if it exists at some time, *regardless* of how those times are related. On this view, past-alteration, like all change, makes no difference to *what there is*, only to *what those things are like*. In particular, it involves turning a bunch of former-events (like *Terrorist Attack*) into non-former-events, and a bunch of non-former-events (like *Guitar Solo*) into former-events. According to the second view, things like *Guitar Solo* only exist at times where they either have occurred, will occur, or are occurring. On this view, past-alteration, *unlike* other sorts of change, *does* make a difference to what there is. In particular, it involves *taking out of* existence a bunch of former-events (such as *Terrorist Attack*) and *bringing into* existence a different bunch of former-events (such as *Guitar Solo*).

We've seen in this section that how one interprets models like $\mathbf{M_2}$ will depend on one's background metaphysics of time and change. Nevertheless, I won't take a stand here on which metaphysics of time is correct. Indeed, the main point I want to emphasise is that *whatever* one's background metaphysical theory, models like $\mathbf{M_2}$ represent genuine change in what is past, as that background metaphysical theory conceives of it. Unlike some other approaches, then, my approach to understanding past-alteration does not presuppose either the A-theory, or the B-theory, or particular versions of these views.[16] As long as one is willing to contemplate the possibility of time being branching in structure and succession coming apart from precedence, then one has the resources to make sense of past-alteration, regardless of one's views about what times are, how objects persist, and what it is for things to change.

6. Hypertemporal Accounts

I am not the first in the literature to suggest that it might, after all, be metaphysically possible to change the past. In particular, several authors (Meiland 1974; Goddu 2003, 2011; van Inwagen 2010; Hudson and Wasserman 2010; Bernstein 2017) have argued that we could make sense of past-alteration if there was an extra dimension of time relative to which we could sensibly talk of *change in what is true at a time*.

To illustrate, consider the following (lightly butchered) analogy due to Goddu (2003). Imagine I insert DVD of *The Lion King* into a (rewritable) DVD player. There is a display on the machine indicating how much time has elapsed on the DVD being viewed – call this 'disk-

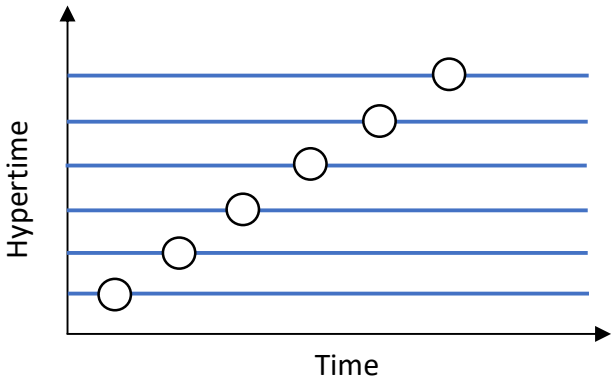
[16] Wasserman's account, for example, "*presupposes the A-theory of time*" (Wasserman 2017: 106); van Inwagen's account "presupposes the 'growing block' theory of time" (van Inwagen 2010: 6); Loss's account "assume[s] exdurantism as the correct theory of persistence (2015: 6); and Goddu's (2003, 2011) account is "seeking to understand time travel within a B-theoretic framework", albeit a two-dimensionalist one (at least according to Baron (2015: 132)).

time'. There is also a digital clock on the wall measuring ordinary external time. I press play on the DVD player at midnight, so that disk-time starts out synchronised with external time. At half-midnight, however, I rewind the DVD to the beginning and, five minutes later, start recording the latest episode of *Love Island: The Unseen Bits*. After ten minutes of this, at 00:45 external time, I stop recording. The data on the disk during the first ten minutes of disk-time is now different to what it was 15 minutes ago – there used to be an uplifting Elton John song at 00:05 disk-time, and now there's footage of reality TV contestants making pancakes. But there's no contradiction here – what data the disk contains at a given disk-time can be different at different external times.

According to proponents of the hypertime account, something similar is going on in cases of past-alteration. To make sense of these cases, they argue, we need to imagine that there are not one but *two* temporal dimensions – time and *hyper*time – such that what is the case at a time can be different at different hypertimes. Proponents of this approach don't usually provide alternative tense logics to accompany their view, but here is a sketch of what such a logic might look like. First, we extend TL by adding two hyper-tense operators, $\mathscr{P}$ and $\mathscr{F}$ – call the resulting language $TL^{(+)}$. As before, several other operators can be defined in terms of $\mathscr{P}$ and $\mathscr{F}$; so, for example, $\mathscr{G}\varphi =_{df} \neg\mathscr{F}\neg\varphi$. A model of $TL^{(+)}$ is an ordered *sextuple*, $M = \langle T, H, \prec, \lhd, V, now\rangle$, where T and H are non-empty sets (which can informally be thought of as the set of times and *hyper*times, respectively), $\prec$ and $\lhd$ are binary relations over T and H, respectively (these can informally be thought of as the relations of precedence and *hyper*-precedence), V is a function which assigns a subset of $\{<t, h>: t \in T, h \in H\}$ to every atomic formula of $TL^{(+)}$ (informally, the set of pairs $<t, h>$ at which that atomic formula is true), and *now* is a function which assigns to every element of H an element of T (informally, the time that is *present* at that hypertime). The truth of a formula φ of $TL^{(+)}$ in model M at time t and hypertime h is then defined as follows:

- If φ is an atomic formula of $TL^{(+)}$, $M, t, h \vDash \varphi$ iff $<t, h> \in V(\varphi)$.
- $M, t, h \vDash \mathbf{P}\varphi$ iff $M, t`, h \vDash \varphi$ for some $t`$ such that $t` \prec t$.
- $M, t, h \vDash \mathbf{F}\varphi$ iff $M, t`, h \vDash \varphi$ for some $t`$ such that $t \prec t`$.
- $M, t, h \vDash \mathscr{P}\varphi$ iff $M, now(h`), h` \vDash \varphi$ for some $h`$ such that $h` \lhd h$.
- $M, t, h \vDash \mathscr{F}\varphi$ iff $M, now(h`), h` \vDash \varphi$ for some $h`$ such that $h \lhd h`$.

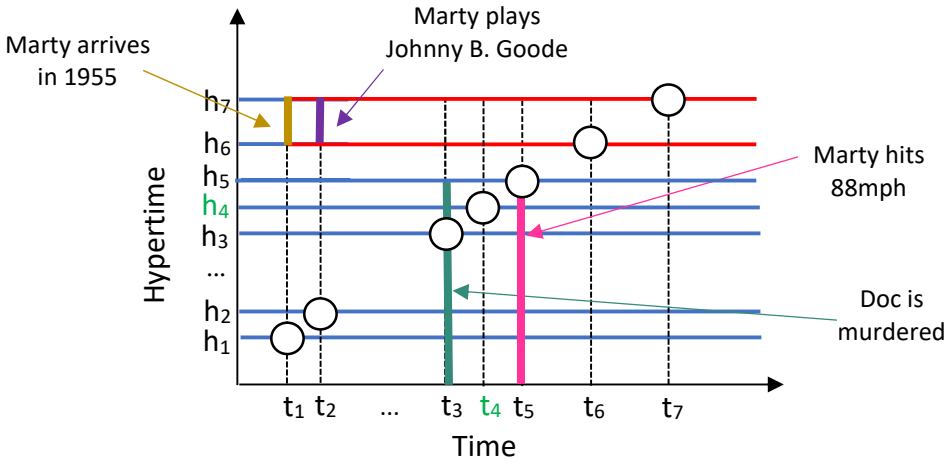
To see how $TL^{(+)}$ might apply to *Back to the Future*, consider first the following visualisation of an 'ordinary' $TL^{(+)}$ model (i.e. one in which no time-travel occurs):



M4

As we move up the hypertime dimension in **M4**, successive times acquire the property of *being present*, as represented by the white circles. But other than that, what is true at a particular time is the same at every hypertime. In models like **M4**, so long as φ doesn't make reference to the property of being present, **P**φ iff $\mathscr{P}\varphi$, and **F**φ iff $\mathscr{F}\varphi$. (Roughly, this is because any time which can be reached by moving right/left along the time dimension can also be reached by moving up/down along the hypertime dimension and across to the white circle.)

Things get more complicated in cases of time-travel, however. There are actually two rival hyper-temporal accounts of past-alteration in the literature. According to the first view,[17] *Back to the Future* is best represented by the following model:



M5

Here t_1 is 06.38 on 5th November 1955 (the time Marty travels back to), and t_2 is later that same day. t_5 is the time Marty travels back from. The current time and hypertime are t_4 and h_4,

[17] This is the view described in Hudson and Wasserman (2010).

respectively. The temporal co-ordinates of important events mentioned in *Back to the Future* are labelled. At hypertimes h_1–h_5, Marty doesn't exist at t_1 and t_2, Doc is murdered at t_3, and Marty disappears after hitting 88mph at t_5. At hypertimes h_6 onwards, however, Marty pops into existence at t_1, plays Johnny B. Goode at t_2, and Doc isn't murdered at t_3 (having been warned by Marty about the Libyans).

It might not be clear, on the face of it, how this model could get us the results we want in *Back to the Future*. After all, (GP) and (LIN-P), the principles apparently violated in *Back to the Future*, are *both true at all times and hypertimes* in $\mathbf{M_5}$, as indeed are their hyper-temporal counterparts. But the hypertime view has a trick up its sleeve here. Once we introduce hypertime, there are two different things one might mean by English sentences like 'It was the case that φ' – either $\mathbf{P}\varphi$ or $\mathcal{P}\varphi$. In normal cases we don't need to be sensitive to this distinction, since $\mathbf{P}\varphi$ iff $\mathcal{P}\varphi$ (for all φ except those that mention the property of presentness). In models like $\mathbf{M_5}$, however, this is no longer the case. If d is the sentence 'Doc is murdered by terrorists', for example, then at the temporal co-ordinate $<t_6, h_6>$, $\mathbf{P}d$ is false in $\mathbf{M_5}$ (because there is a no time $t \prec t_6$ along the h_6 axis at which Doc is murdered) but $\mathcal{P}d$ is true (because there is a hypertime $h \lhd h_6$ such that Doc is murdered at $<now(h), h>$, namely h_3).

Now consider again the sentences that are seemingly entailed by *Back to the Future*, starting with (1):

(1) Marty has never played and will never play Johnny B. Goode at a high-school prom, but it will be the case that he did.

(1) has several interpretations in $\mathrm{TL}^{(+)}$, depending on whether the English tense operators are interpreted as expressing the ordinary Priorian tense operators or the hypertense operators. As before, let p stand for the sentence 'Marty is playing Johnny B. Goode at a high-school prom'. The first conjunct has two interpretations: $\neg\mathbf{S}p$ and $\neg\mathcal{S}p$. As it happens, both of these are true in $\mathbf{M_5}$ at $<t_4, h_4>$ (the time and hypertime at which *Back to the Future* is set) – the former is true because p is false at all times on the h_4 axis, and the latter is true because p is false at $<now(h), h>$ for all h (i.e. at all the white circles). The more interesting conjunct is the second, namely 'It will soon be the case that Marty played Johnny B. Goode at a high-school prom'. This has four interpretations: (a) $\mathbf{FP}p$, (b) $\mathcal{F}\mathcal{P}p$, (c) $\mathbf{F}\mathcal{P}p$, and (d) $\mathcal{F}\mathbf{P}p$. (a) is false in $\mathbf{M_5}$ at $<t_4, h_4>$ because there is no time on the h_4 axis at which p is true. (b) is false because p is false at $<now(h), h>$ for all h (i.e. at all the white circles), and (c) is false for the same reason. But (d) is true in $\mathbf{M_5}$ at $<t_4, h_4>$. This is because there is a hypertime h such that $h_4 \lhd h$, and a time $t \prec now(h)$, such that p is true at $<t, h>$ (for example, p is true at $<t_2, h_6>$). By postulating an additional temporal dimension, then, the proponent of the hypertime view can coherently

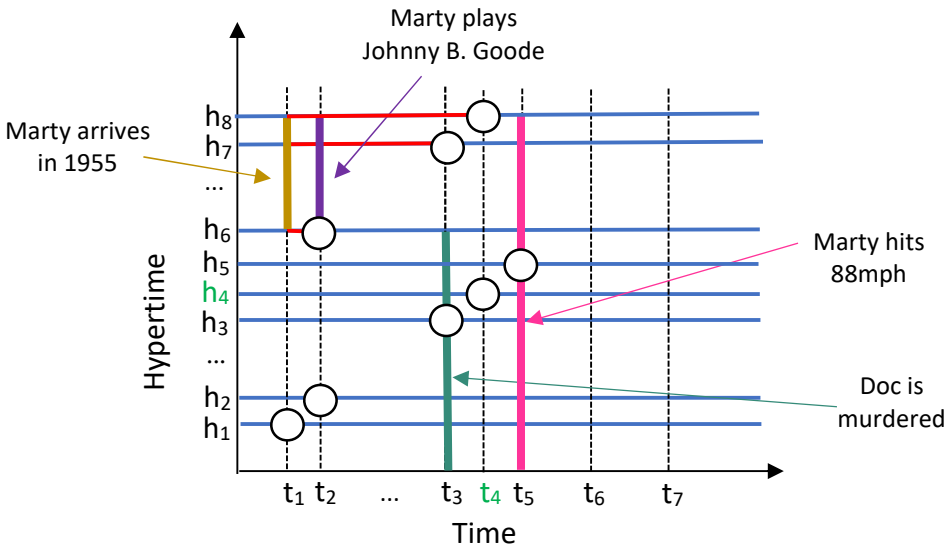
maintain that there is a true interpretation of claims like (1) in *Back to the Future*, even though all instances of (LIN-P) and its hyper-temporal analogue are true in $\mathbf{M_5}$.

The same is true of (2):

(2) Doc Brown's body is ridden with bullets, but it will be the case that Doc's body has never been ridden with bullets.

As before let q stand for the sentence 'Doc's body is ridden with bullets'. The first conjunct of (2) is straightforwardly true at $<t_4, h_4>$. The second again has several interpretations, namely $\mathbf{FH}\neg q$, ~~$\mathbf{FF}$~~$\neg q$, $\mathbf{F}$~~$\mathbf{F}$~~$\neg q$, and $\mathcal{F}\mathbf{H}\neg q$. The first three are false in $\mathbf{M_5}$ at $<t_4, h_4>$, but the fourth is true, since there is a hypertime h such that $h_4 \lhd h$ (for example, h_6) such that at no time along h's axis is Doc's body ridden with bullets. So the proponent of the hypertime view can again coherently maintain that there is a true interpretation of claims like (2) in *Back to the Future*, even though all instances of (GP) and its hyper-temporal analogue are true in $\mathbf{M_5}$.

On the second version of the hypertime view – the 'moveable present' view[18] – *Back to the Future* is better represented by means of the following model:



$\mathbf{M_6}$

In $\mathbf{M_6}$, the present moment travels back *with* Marty from 1985 to 1955, and the changes to what is the case from t_1 onwards happen, not all at once, but gradually as hypertime evolves. For example, at h_6 Marty exists at t_1 but goes out of existence at t_2; by h_7, Marty is playing Johnny B. Goode at t_2, but Doc is still murdered at t_3; by h_8, Doc is alive at t_3. (I'm assuming here that the present moment travels instantaneously back to t_1 as soon as Marty hits 88mph,

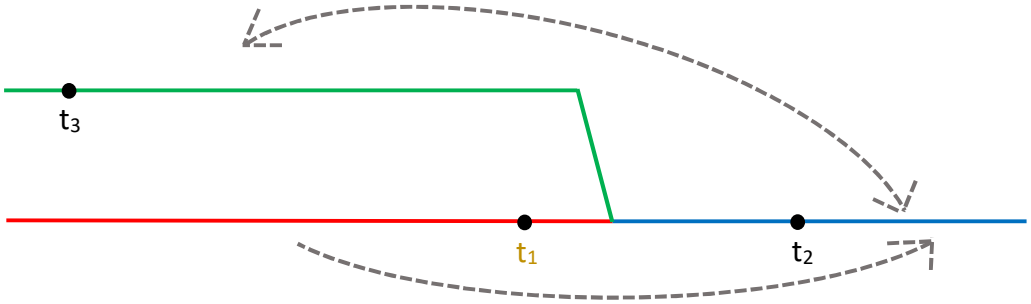
[18] For an explanation and defence of the moveable present view, see Bernstein (2017).

but this isn't required – $TL^{(+)}$ allows for models where the present takes a certain interval of hypertime to travel back, just like the DVD takes a certain interval of external time to rewind.)

$\mathbf{M_6}$, just like $\mathbf{M_5}$, allows for true interpretations of (1) and (2). But there are also some interesting differences between the two models. Recall that, in $\mathbf{M_5}$, both interpretations of 'Marty has never and will never play Johnny B. Goode at a high-school prom' – namely $\neg\mathbf{S}p$ and $\neg\overleftrightarrow{\mathbf{S}}p$ – are true at $<t_4, h_4>$. In $\mathbf{M_6}$, however, the former interpretation is true but the latter is false – there *is* a hypertime h such that Marty is playing Johnny B. Goode at $<now(h), h>$, namely h_6. According to $\mathbf{M_6}$, then, there is a sense in which Marty will be playing Johnny B. Goode at a high-school prom in *Back to the Future*, as well as a (different) sense in which he will never do so.

I mention these differences between the two models only to illustrate how their predictions come apart; I'll leave it up to proponents of the hypertime account to debate their relative benefits. I also won't get into the difficult question of whether the situations represented by models like $\mathbf{M_5}$ and $\mathbf{M_6}$ are metaphysically possible, and in particular, whether it's metaphysically possible for there to be more than one dimension of time. Instead I want to focus on what I think is a more fundamental issue with the hypertime accounts of past-alteration. Recall that what we wanted was a way of making sense of change (in the ordinary sense) in what is past (in the ordinary sense). But this is not, in fact, what the hypertime accounts give us. Instead what they give us is models in which there is *hyper*-change in what is past, where hyper-change stands to ordinary change as hypertime stands to ordinary time. The question of whether there could be hyper-change in what is past might, of course, be an independently interesting metaphysical question. But it is not, I think, the question we started out with. The question we started out with was *not* whether there could have been an additional dimension along which there is variation in what is true at a time; it was whether there could be *change* in what is *past*, with 'change' and 'past' interpreted as they are standardly interpreted, relative to the same dimension (namely, *time*).

Now, proponents of the hypertime account may well protest that if 'change' and 'past' are interpreted relative to the same dimension, then change in what is past would necessarily require there to be a difference in what is true at a time *at different times*, and this is straightforwardly incoherent. But in fact, as I showed above, past-alteration does *not* require change, in *any* sense, in what is true at particular times. Consider models like $\mathbf{M_2}$ again:



M₂

In **M₂**, there is no 'change' in what is true at t_1, or t_3, or any other time; what changes is rather *which of those times are in the past*. Before Marty hits 88mph in *Back to the Future*, the times 30 years in his past are ones in which he isn't alive; after Marty hits 88mph, the times 30 years in his past are ones in which he's playing Johnny B. Goode at a high-school prom. There is no contradiction here, because the times precedence-related to t_1 are different to those precedence-related to t_2. And to repeat, this is not some *ad hoc* stipulation – it falls very naturally out of attempts to construct temporal models on which past-alteration scenarios come out true, models in which the principles apparently violated in past-alteration scenarios, namely (LIN-P) and (GP), are falsified, not just on *some* interpretations, but *simpliciter*.

In summary, the received wisdom is mistaken in thinking that past-alteration requires hypertime. What matters for whether past-alteration is possible is not whether there could have been an additional dimension of time, but rather whether the single dimension of time that actually exists could have had a particular kind of *structure*, the kind of structure represented by models like **M₂**.

7. Past-Alteration and Explanatory Gaps

So far, my focus has been on explaining how there could be change in what is past. But there is another important feature of cases like *Back to the Future* that needs to be considered. Marty wasn't alive 30 years ago in *Back to the Future*; after he hits 88mph in the DeLorean, it will be the case that he was. But there is also a *causal* connection between these facts. It is no mere coincidence that the past will be different from how it is now after Marty hits 88mph; it's *because* Marty is about to hit 88mph that the past is about to change. In particular, the following seems true in *Back to the Future*:

(5) It is because Marty is about to hit 88mph in the DeLorean that he will soon have been playing Johnny B. Goode at a high-school prom.

The question is whether we can adequately capture (5) on the model of past-alteration I have been advocating.[19]

Here's one reason for doubting that (5) can be adequately captured. Causation, according to many philosophers, is a *relation* between two events or states of affairs. But at t_1, the time *Back to the Future* is set, there is no past (or future) event of Marty's playing Johnny B. Goode at a high-school prom – the past, after all, has not yet been altered. How, then, could (5) possibly be grounded in a relation holding between two things, if one of them never has occurred and will never occur?

Now, as discussed in §5, it doesn't necessarily follow from the fact that *Guitar Solo* (the thing which at t_3 is the event of Marty's playing Johnny B. Goode) has never occurred and will never occur at t_1 that it doesn't *exist* at t_1. Assuming it does exist at t_1, it is not a former-event at t_1, since it never has been an event; nor is it a future-event, since it never will be an event. But it is a *future-former-event* – something which *will have been* an event. It would therefore be open to a proponent of this view to ground the truth of (5) at t_1 in a relation that holds between two, existing things: a future-event (the thing which will be Marty's hitting 88mph) and a future-former-event (*Guitar Solo*). But the idea that the causal relation could hold between two things neither one of which precedes the other is, to say the least, very strange indeed. Philosophers disagree, of course, about whether causes must precede their effects. But as far as I know, no-one has taken seriously the idea that a cause could occur neither before, nor after, nor at the same time as its effect!

Perhaps a better strategy, then, might be to simply deny that facts like (5) need be grounded in a relation between two events. It's long been recognised that presentists face a problem accounting for even simple causal claims like 'The short circuit caused the fire' in terms of a relation between two things, since on their view there is no time at which the short circuit and the fire both exist. In response, these philosophers have mostly argued that causation shouldn't be thought of as a relation between events after all.[20] Instead, basic causal claims are simply claims of the form 'φ BECAUSE ψ', where 'BECAUSE' is a special kind of sentential connective.[21] Claims that seem to predicate a causal relation between events, like 'The short circuit caused the fire', should really be understood as paraphrases of claims of the form 'A fire

[19] Wasserman (2017: 104-5) briefly raises what I think is a similar challenge, although his objection is directed much more narrowly at attempts to make sense of past-alteration that combine the A-theory with eternalism.

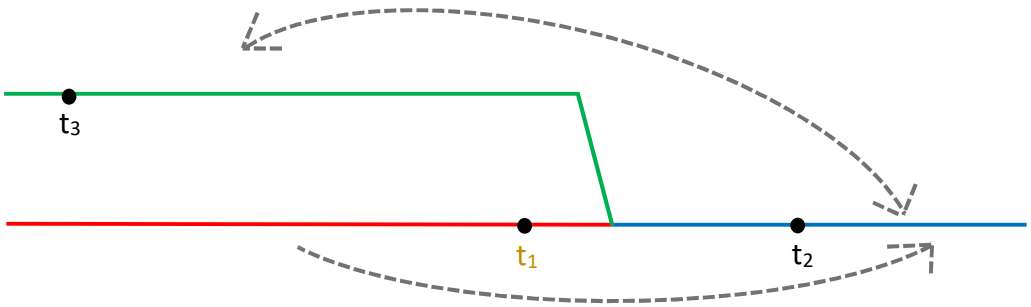
[20] See Sider (1999); c.f. Markosian (2004).

[21] It's possible to think of 'BECAUSE' as referring to a relation between *facts*, but one needn't reify things in this way, and indeed many proponents of this view explicitly deny that causation is any kind of relation at all.

broke out BECAUSE a short circuit occurred', which aren't committed to the existence of non-present events. Similarly, on this view, we can truly say that Marty will have been playing Johnny B. Goode at a high-school prom *because* he is about to hit 88mph in the DeLorean, without committing ourselves to the present existence of something which will have been Marty's playing Johnny B. Goode.

Applying this strategy to *Back to the Future* reveals an interesting difference between everyday causal claims and the sorts of causal claims that arise in past-alteration scenarios. In most everyday cases, *iterated* tensed claims like **FP***p* plausibly don't admit of direct causal explanations; rather, their truth is simply *grounded* in the truth of non-iterated tensed claims like **P***p* (which may themselves, of course, have causal explanations). The fact that it will be the case tomorrow that France won the 1998 FIFA World Cup, for example, is true simply in virtue of the fact that France won the 1998 FIFA World Cup in 1998. Interestingly, however, this is not the case in *Back to the Future*. That Marty will soon have been alive 30 years ago is *not* true in virtue of the fact that Marty was alive 30 years ago, since he wasn't; rather, the fact that Marty will soon have been alive 30 years ago is *directly* causally explained by the fact that he will soon hit 88mph in the DeLorean. This is precisely what is distinctive about past-alteration: genuinely altering the past involves more than just doing something that causally explains why the past is as it is; it involves doing something that causally explains why the past *will* soon be *different* from how it now is.

Unfortunately, though, things get more complicated when we consider what *will* be true in *Back to the Future*, after Marty hits 88mph. Here again is my TL$^{(-)}$ model of the case:



M₂

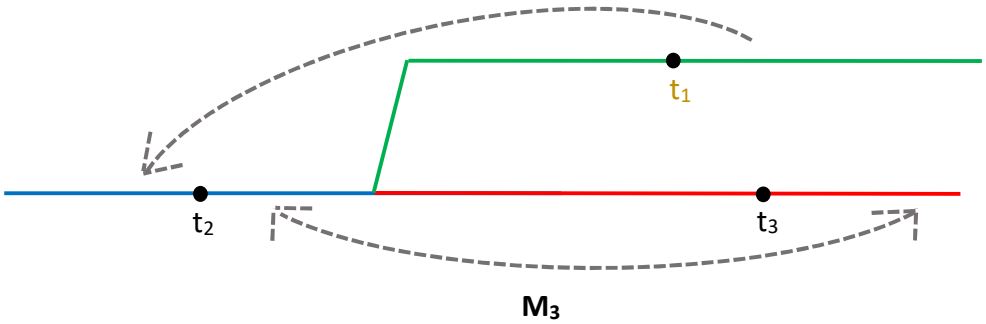
Consider what is true at t_2 in **M₂**:

Back to the Future (Future)

It's 1.30am on the 26th October 1985. Thirty years ago, something strange happened: a teenage boy appeared out of nowhere in a DeLorean car, claiming to have memories of all kinds of things – the death of his friend Doc Brown at the hands of Libyan terrorists, for example – that have never happened nor will ever happen.

Here's the problem. In *Back to the Future (Future)*, it's the case that Marty appeared out of nowhere 30 years ago (i.e. at t_3) with a bunch of memories of things that, from the perspective of t_2, never happened and will never happen. But what, if anything, causally explains this fact? The description of *Back to the Future* leaves open what exactly happens after t_3 on the green branch; in particular, it leaves open whether Marty hits 88mph in a DeLorean at any point.[22] But even if he did, it's not at all clear that this would explain why he turned up 30 years ago with a bunch of apparent memories of things which never happened. So even though, from the perspective of t_1, we can truly say that Marty will have appeared 30 years ago because he is about to hit 88mph in the DeLorean and travel back in time, it nevertheless seems that from the perspective of t_2, it is *inexplicable* why Marty appeared 30 years ago, because the fact which at t_1 explains why he will have appeared 30 years ago is no longer be the case after he hits 88mph. In other words, past-alteration creates *explanatory gaps*: by changing the past, we often also change that which now explains why the past will soon be different.

The problem here is in some ways even more vivid in cases of future-alteration, like *Back to the Future II*. Recall that in *Back to the Future II*, Marty Jr. was *going* to be convicted of robbery, but ever since Marty and Doc hit 88mph in the DeLorean in 1985 he has never and will never be convicted of robbery. Here again is my TL$^{(-)}$ model of this case:



[22] As it happens, in the 'Back to the Future' movie Marty does indeed hit 88mph in the DeLorean at the last instant of the 'altered' past, watched by himself; but this certainly isn't required to make the story consistent.

As with *Back to the Future*, there seems to be some important explanatory connection between the fact that Marty Jr. won't be convicted of robbery and the fact that Marty and Doc hit 88mph in 1985 – it's *because* Marty and Doc did what they did 30 years ago that Marty Jr. now won't go to prison. But now consider the same case from the point of view of t_2, before the future was altered:

Back to the Future II (Past)
It's the 26th October 1985. Doc and Marty are about to hit 88mph in the DeLorean. They intend to travel to 2015 to prevent Griff from meeting Marty Jr. and ruining his life. But they will fail – Marty Jr. will be coaxed by Griff into committing a robbery, caught by the police, and sentenced to 15 years in prison.

It's part of the set-up of *Back to the Future II* that at t_2, it's the case that Marty Jr. will be convicted of robbery; *a fortiori* it's part of the set-up of *Back to the Future II* that, at t_2, any attempts to prevent this will fail. Moreover, Doc presumably knows this (it's only because of his knowledge of the future that he intends to change it). From the perspective of t_2, then, it seems hard to explain why Doc is going to so much trouble to try to prevent something he already knows will happen. Even though, from the perspective of t_1 (the time *Back to the Future II* is set), we can truly say that Doc's attempt to change the future was successful – the future is indeed different now to how it was going to be – nevertheless from the perspective of t_2, Doc's future actions seem inexplicable, since he knows (or at least should know) that Marty Jr. will be convicted of robbery despite his actions. So future-alteration also seems to involve explanatory gaps: any evidence which could motivate an attempt to change the future is necessarily evidence that it will fail, and so any such attempts must seemingly be irrational, even if with the benefit of hindsight we can truly say that past such attempts were successful.

What we've revealed in this section is yet another way in which past- and future-alteration scenarios confound our usual expectations about what can change over time. Usually, if it's explicable *now* why P will be true at some future time, it will still be explicable, at that future time, why P is true at that time; *not so* in past-alteration cases, where by altering the past we can also alter what now explains why the past will soon be different. Usually, if an action at a past time is rationalizable now, then it was also rationalizable at that past time; *not so* in future-alteration cases, where the reasons which now make an action rational only came into existence after, and indeed *because*, that very action was performed. One might well wish to leverage these results into an argument that past- and future-alteration scenarios, no matter how coherent they might appear at first, are indeed impossible, on the grounds that the theoretical costs of admitting their possibility are simply not worth the benefits. But I will not come down on either side of this debate. My goal in this paper has merely been to arrive at a

richer understanding of what the world would have to be like for past- and future-alteration to occur; I will leave it to others to judge whether the world could, in fact, have been that way.

8. Conclusion

In this paper, I have sought to tackle head-on the commonly expressed, but rarely defended, accusation that past-alteration is 'logically impossible'. I started by showing that past-alteration cases violate two important tense-logical principles: (LIN-P) and (GP). I then sketched a novel tense logic whose models allow for false instances of these principles, before discussing in detail what a world accurately described by such models would actually be like. Against the received philosophical wisdom, I argued that such a world would *not* be one in which there are two dimensions of time, but rather one in which time has a certain kind of *structure*: one which is 'backwards-branching', and where the relation of succession 'comes apart' from the relation of precedence. Whether time could indeed have had that structure is a metaphysical question, not a logical one. Nevertheless, I hope to have demonstrated how tense logic can play an important role in helping us to get clear on what is at stake in the debate over whether past-alteration is possible.

References

Baron, S. (2017). Back to the Unchanging Past. *Pacific Philosophical Quarterly 98*(1), 129-147.

Bernstein, S. (2017). Time Travel and the Movable Present. In J. Keller (ed.), *Being, Freedom, and Method*: *Themes from the Philosophy of Peter van Inwagen*, Oxford: OUP.

Brier, B. (1973). Magicians, Alarm Clocks, and Backward Causation. *Southern Journal of Philosophy 11*(4), 359-364.

Brown, G. (1985). Praying about the past. *Philosophical Quarterly 35*, 83-86.

Deasy, D. (forthcoming). Advanced Temporalising. *Oxford Studies in Metaphysics, Volume 12*. Oxford: OUP.

Dwyer, L. (1977). How to Affect, but not Change, the Past. *The Southern Journal of Philosophy*, *15*(3), 383-385.

Flew, A. (1954). Can an Effect Precede its Cause? *Proceedings of the Aristotelian Society 28*, 45-62.

Fulmer, G. (1983). Cosmological Implications of Time Travel. In R. E. Myers (ed.), *The Intersection of Science Fiction and Philosophy*, Westport, CT: Greenwood.

Gaspar, E. (1887 [2012]). *The Time Ship: A Chrononautical Journey* (trans. Yolanda Molina-Gavilán and Andrea Bell). Middletown, CT: Wesleyan University Press.

Goddu, G. C. (2003). Time Travel and Changing the Past: (Or How to Kill Yourself and Live to Tell the Tale). *Ratio 16*(1), 16-32.

Goddu, G. C. (2011). Avoiding or Changing the Past? *Pacific Philosophical Quarterly 92*, 11-17.

Goranko, V. and Galton, A. (2015). Temporal Logic. In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy (Winter 2015 Edition)*, URL = <https://plato.stanford.edu/archives/win2015/entries/logic-temporal/>.

Horwich, P. (1975). On Some Alleged Paradoxes of Time Travel. *The Journal of Philosophy 72*(14), 432-444.

Hospers, J. (1967). *An Introduction to Philosophical Analysis*. London: Routledge & Kegan Paul.

Hudson, H. and Wasserman, R. (2010). Van Inwagen on Time Travel and Changing the Past. In D. Zimmerman (ed.), *Oxford Studies in Metaphysics, Volume 5*, Oxford: OUP.

Lewis, D. (1976). The Paradoxes of Time Travel. *American Philosophical Quarterly 13*(2), 145-152.

Loss, R. (2015). How to Change the Past in One-Dimensional Time. *Pacific Philosophical Quarterly 96*(1), 1-11.

Markosian, N. (2004). A Defense of Presentism. In D. Zimmerman and K. Bennet (eds.), *Oxford Studies in Metaphysics, Volume 1*, Oxford: Oxford University Press.

Prior, A. N. (1957). *Time and Modality*. Oxford: Clarendon Press.

Prior, A. N. (1967). *Past, Present and Future*. Oxford: Clarendon Press.

Prior, A. N. (1969). *Papers on Time and Tense*. Oxford: Clarendon Press.

Sider, T. (1999). Presentism and Ontological Commitment. *The Journal of Philosophy 96*(7), 325-347.

Sider, T. (2001). *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: OUP.

Sider, T. (2002). Time Travel, Coincidences and Counterfactuals. *Philosophical Studies 110*(2), 115-138.

Sider, T. (2006). Quantifiers and Temporal Ontology. *Mind 115*(457), 75-97.

Smith, N. J. J. (1997). Bananas Enough for Time Travel? *British Journal for the Philosophy of Science 48*, 363-389.

van Inwagen, P. (2010). Changing the Past. In D. Zimmerman (ed.), *Oxford Studies in Metaphysics, Volume 5*, Oxford: Oxford University Press.

Vihvelin, K. (1996). What Time Travelers Cannot Do. *Philosophical Studies 81*(2-3), 315-330.

Vranas, P. B. (2005). Do Cry Over Spilt Milk: Possibly You can Change the Past. *The Monist*, *88*(3), 370-387.

Vranas, P. B. (2010). What Time Travelers May be Able to Do. *Philosophical Studies 150*(1), 115-121.

Wasserman, R. (2017). *Paradoxes of Time Travel*. Oxford: OUP.