

On the Compatibility of Free Will and a Fixed Future

Matthew Evpak

California State University, Sacramento

Abstract: It is possible for human free will to exist even if the future is fixed. To say that the future is 'fixed' is to say that propositions concerning the future, including propositions describing the future actions of humans, are true in the present and have always been true. The sort of compatibility advanced in this account relies on a particular metaphysical view called the four-dimensional theory of time, which maintains that time is to be understood perspectively: 'past' and 'future' are simply relations between an observer and a point in time, never intrinsic properties of points in time.

There is a common notion of time and human action according to which persons may act freely only if the future is in some sense 'open' or indeterminate. Free will is held to rely conceptually on there being no definite future state of affairs, because a (fully) predetermined future is one which is beyond a person's power to alter. What a person does cannot be done freely if it was already fixed that he would do so – in such a case, it seems, the person could not act otherwise, and so is constrained to act only as he in fact does. The compatibility issue implicit here might come to the fore in light of logical determinism, the theory that the future is fixed in virtue of the unalterable truth or falsity of propositions concerning the future; or it might arise in a theological context regarding human free will vis-à-vis incorrigible divine foreknowledge.¹ Whatever the initial cause of inquiry, many have an intuition that it is a necessary condition of human free will that the future be indeterminate and at least partially governable for humans.² If we wish to maintain a view of human existence which affords free choice to human action, it seems we are obliged to deny categorically any incorporation of a fixed future into our ontology. Yet I am not convinced that the conflict here is so severe, if indeed there is one, for it seems to me that free will and a fixed future are quite capable of belonging to a single coherent metaphysical scheme. Although it is the determinacy of the future that seems to preclude human free will, I think that, contrary to the popular notion, the determinacy of the future is an

integral part of a reconciliatory account of time and human action. To be specific, I am here proposing that a particular theory of time – which I shall call the ‘four-dimensional’ theory of time – yields a compatibility between human free will and a fixed future.

It will be helpful to introduce and clarify a number of terms. When I speak of a ‘four-dimensional’ theory of time, I mean the idea that time is a dimension whose characteristics are analogous to spatial characteristics³; for instance, just as one’s spatial location is a product of one’s orientation relative to other objects in space, one’s temporal location is a product of one’s orientation relative to events (which might be thought of as the analogue of spatial objects: an event is a ‘temporal object,’ as it were). In this way, the four-dimensional theory of time is at its core a *perspectival* theory: past, present, and future are not absolute ontological demarcations but simply artifacts of an observer’s vantage point, much as left, right, up, down, ‘here,’ and ‘there’ are not absolute spatial properties so much as relationships to particular orientations in space. Another way of expressing this is to say that, just as ‘being above’ and ‘being below’ are relational properties of spatial objects rather than essential or intrinsic properties, ‘being past’ and ‘being present’ and ‘being future’ are not properties that times possess in themselves but are merely kinds of relationship that times can bear with respect to a temporal vantage point. The implication of the four-dimensional theory of time is therefore that in whatever sense past, present, and future exist, all three exist in the same way, because all three in themselves are the *same sort* of thing, being distinguished not by any intrinsic nature they may have but only by their relationship to a given perspective.⁴

Fatalism follows naturally from the four-dimensional theory of time. I use ‘fatalism’ to refer to the idea that the future is determinate or (what is the same) ‘closed,’ which notion is itself in need of explication: for any instant or temporal ‘point’ t , t is closed if and only if the truth value of any given proposition concerning the state of affairs of the world at t is the same at all times – or, we might say, is the same *eternally*. In other words, a time t is closed if and only if whatever is true at t of the state of affairs at t is true at all times: ‘now’ is closed iff all currently true propositions concerning the present state of the universe always have been and always will be true. To say that the *future* is closed, then, is to say that for any present time t , all times $t + n$ are closed – all true propositions concerning the future are true *now*, to put the thesis of fatalism informally.

So fatalism, the theory that all future times are closed, is a corollary of the four-dimensional theory of time, because the latter

implies that *all* times are closed (in this sense, fatalism is simply the 'latter half' of the four-dimensional theory). Because of the relativity of past, present, and future, there is a sense in which characteristics of one of the three are generalizable to the others. If it is true to say in the year 2001 that *x* is the case in the year 2000, why should it not be true in the year 1999 to say that *x* is the case in the year 2000? While it is true that propositions about the year 2000 are 'past' in the year 2001 and 'future' in the year 1999, this distinction should have no bearing on what is actually true about the year 2000; it makes little sense to say that the truth value of the proposition differs in virtue of its being past or future if such terms are merely relational and not inherent to the year 2000. According to the four-dimensional theory of time, truth possesses a sort of atemporality or eternality – truth is, as it were, 'outside of time' and so is not determined, altered, or in any way affected by it; and this, in effect, is the thrust of what it means for a time to be 'closed.'

What remains to be defined is free will. For the purposes of this essay, I do not see the need to provide or insist upon a particular interpretation of the concept of free will (nor do I wish to devote too much space to an issue largely beyond the scope of the essay), but I will say that I employ 'free will' in what I take to be a common and intuitive sense: the ability to act without one's actions being wholly caused or determined by anything external to oneself. I realize that free will is a complex and controversial subject and that there is dispute as to whether the very concept of free will is even intelligible, but all I am contending here is that *insofar as* humans may possess free will, such freedom is not negated by the determinacy of the future. It should also be noted that whether free will is compatible with *determinism* is entirely another matter, having (in my estimation) more to do with causality than with time; as such, causal relationships will be of limited applicability to this matter. And with respect to the debates surrounding theories of human action and theories of time, I must note that my purpose is not to argue that human free will in fact exists or that the four-dimensional theory of time (and thus fatalism) is somehow superior to its alternatives, only that the two are consistent with each other.

With this terminology in place, we can begin to examine the ramifications of treating time four-dimensionally, and to that end I think it will prove useful to devise a way of 'visualizing' time, so to speak. Returning to the analogy between space and time, it seems to me that when one is attempting to generate a more objective conceptualization of something perspectival in nature, it helps to 'remove' oneself mentally from the thing by introduction of a new dimension. So in order to visualize a plane, one might add a third

dimension and imagine the plane from 'above' it rather than from a point *on* it. Similarly, in order to visualize time, which humans experience one-dimensionally and thus linearly, I suggest we add a second dimension and view the line from a point not on it.⁵ Indeed, this is precisely the purpose of a *timeline* – it furnishes the viewer with a means of visualizing time 'from without,' a way to think of temporal objects *apart* from their relationship to an observer in time. And this is also the essence of the four-dimensional account: without any particular vantage point within time (or on the timeline, one might say), the distinction among past, present, and future simply *vanishes*. On a timeline there *is* no past, present, or future. Someone looking at the timeline will probably interpret it with respect to her own temporal circumstance, but this is no feature of the timeline *itself*, as someone who looks at the timeline centuries later will interpret it quite differently in this regard. Thus any assignment of 'past,' 'present,' or 'future' to points on a timeline is precisely that – an artificial imposition of values onto times which lack any such values otherwise.⁶

I propose that there is exactly one timeline corresponding to the actual world (given a frame of reference⁷); that is, assuming that a timeline can in principle convey an infinite amount of information and completely describe every instantaneous state of affairs in the actual world,⁸ there is only one that describes the actual world and not a merely possible one. This follows from the (I think generally agreeable) supposition that there is only one true history⁹ or 'course' of the world: for any state of affairs *s* and any time *t*, either *s* obtains at *t* or it does not, and every *t* corresponds to exactly one *s* (although *s* may be conceptualized or interpreted in various ways, or not at all). This is the case for any possible world wherein time exists similarly¹⁰ to our world: abstractly speaking, if we take each moment and compress it into a point containing all information regarding the world at that moment, and we arrange the points into a line, we have the spatiotemporal equivalent of a strand of DNA uniquely encoding that possible world. We can then select any point on the line and 'expand' it to ascertain the state of affairs in that possible world at that moment – such data are precisely what a timeline is taken to report. There is *ex hypothesi* a one-to-one correspondence between the set of possible worlds and the set of possible timelines, given that each timeline is 'maximally informed' (reports everything there is to report about its corresponding world).

And so the actual world, being unique, has also a unique timeline documenting its history. Let us refer to the actual world's timeline with the capital letter *T*; the lowercase *t* I have been using can be thought of as a point on *T* or any other possible world's timeline. It is

important to note that *T* spans the entirety of the actual world's temporal existence, from the beginning of time to the end (if the limits exist). All events which transpire throughout the history of the universe are 'written' into *T* timelessly and unchangeably – even those which have not yet happened from a human perspective, for 'not having happened yet' is not a property of a time itself but an artifact of conscious experience. Here, at last, is the apparent conflict between human free will and a fixed future – how can I be free if future events, including my future actions, are already inscribed in *T*, already 'going to happen'? But here as well is the resolution, and the four-dimensional theory of time is the catalyst that brings these inimical elements into synthesis. Here, I believe, is the crux of the entire issue: does human action *obey T* – or *shape* it?

The incompatibilist, as he may be called,¹¹ objects "Surely I cannot have free will, for all of my future actions are already set in stone, and I cannot do otherwise – why, even this very sentence was I bound to utter from eternity!" He gestures to a depiction of *T* whereupon is indicated that he would speak as he just has. But it seems to me that the incompatibilist is here committing a serious and profound error: he has failed to take into account that *T* reports his utterance precisely *because* he utters it. The timeline is as it is *because* of what he does, not the reverse.

Let us consider a thought experiment. Suppose a time traveler named Smith, convinced that a fixed future and human free will are incompatible, decides to strip Jones of his free will. Smith shadows Jones on June 1st, taking detailed notes on Jones' actions, decisions, utterances, and such, and then travels back to May 31st with his notepad. Smith then attends a social gathering (Jones is absent) and recounts for his friends the course of Jones' life on the 'next' day, June 1st. After he has finished reading from his notes, Smith proudly proclaims that he has successfully eradicated any ability on Jones' part to act freely the next day – after all, Smith's notepad will dictate Jones' actions, decisions, and such, and Jones is powerless to resist. "Perhaps it is true after all," says Smith, "that the pen is more powerful than any sword. Behold its coercive might!" But something is clearly amiss in Smith's assessment of the situation; it makes no sense to speak of the notepad or Smith's 'foreknowledge' controlling or constraining Jones' capacity to act on June 1st. In fact, it seems to me that just the opposite is true: Jones' free-willed actions 'fix' Smith's knowledge and notes, and it is *these* which conform to *Jones*. There is a very real sense in which Jones is responsible for the content of Smith's notes and Smith's cognition, as Jones is in fact the one who exerts a sort of creative or determinative power in relation to that content.¹²

And so it is with human action and *T*: even though *T* bears an exhaustive and accurate list of every action a person takes, this does not impose restrictions on a person's behavior, because the content of *T* is, as far as it concerns the person, a *product* of her behavior. People literally 'make history,' or at least contribute to it. And since past, present, and future are fundamentally perspectival attributes, it makes no sense to ask whether an event appears on *T* 'before' it happens, or whether a point on *T* remains vague and indeterminate 'until' it actually comes to pass. Recall that once one has stepped 'outside' time in order to view *T*, there is no 'future' to be indeterminate; such questions simply become meaningless from such a (hypothetical and abstract) vantage point. The complete history of the actual world is laid out simultaneously, and some of the history is as it is solely in virtue of whatever free-willed beings exist in the world.

I acknowledge that this point is disputable: does a history determine a timeline, or does a timeline determine a history? Which way does the 'fixation' work? Admittedly, I have no direct evidence one way or the other, and I am uncertain how one might go about acquiring such evidence, but I find the former stance both more intuitive¹³ and more parsimonious. (What determines the timeline, if not human action and other historical fact? Or are timelines by nature 'brute'?) This is certainly not in any proper sense an *argument* for historical primacy or 'primacy of the event,'¹⁴ but just a starting point for deciding which view to adopt; but either way, my purpose has been only to establish the *compatibility* of free will and fatalism, not the facticity of either. And if it is merely *possible* that free will have primacy over *T*, the disputability of this point is irrelevant to my thesis; as it turns out, I perceive no incoherency in the notion of the more intuitive primacy and therefore no justification for denying its possibility.

One essential point remains to be examined: even if people have a part in shaping *T*, are they not still *bound* by it just the same? The incompatibilist may protest, "Yes, it is all well and good to say that I am the one who predetermines my future actions, but the fact remains that my future actions *are* predetermined, and so I am no less bound to obey that predetermination than if someone else had predetermined my future for me – even if I extinguish my own freedom, the simple fact of the matter is that my freedom *is* extinct!" While this is an insightful and principled objection, it seems to me once again that the incompatibilist has not looked deeply enough into the implications of the four-dimensional account of time; specifically, there are some important points to be made regarding

questions of modality – that is, regarding the possibility or necessity of particular states of affairs.

Let us suppose that on June 1st, Jones has a choice between walking to work or taking the bus. We can express the truth that there is a possible world where Jones walks on June 1st and a possible world where he takes the bus by saying $\diamond W$ or $\diamond B$, respectively. However, he may only choose one or the other, and we can express this exclusivity with $\sim \diamond (W \& B)$ or $\Box \sim (W \& B)$: 'it is not possible that Jones does both' or 'it is necessary that Jones does not do both.' There is a sense in which, in all possible worlds where Jones walks to work, Jones is *bound* to walk to work. This is merely the truism that in all possible worlds, if Jones walks to work, then Jones walks to work: $\Box (W \rightarrow W)$. If in fact it is not the case that Jones walks to work, then we are no longer speaking of possible worlds in which Jones walks to work, for clearly it is not possible for Jones to walk to work and not walk to work: $\sim \diamond (W \& \sim W)$.¹⁵ But I contend that this 'boundness' is best characterized not as necessity, but as what I shall call *inevitability*, a distinctive property of events on the four-dimensional account of time.

To understand inevitability, let us adapt the classic sea battle argument.¹⁶ On the four-dimensional theory of time, at least, it is either true or false *now* that there will be a sea battle tomorrow. That fact is easy enough to accept, but its ramifications may prove surprising: if it is true now that there will be a sea battle tomorrow, then all possible worlds that could be subsequent to the present state of affairs *must* include a sea battle tomorrow. This, it will be seen, is a sort of limited necessity – it is *logically impossible* for it to be true at t that a battle will occur at t' and for there not to be a battle at t' . Once it is true that the battle will take place at t' , all possible futures must include that battle, as those that do not would bring about a contradictory state of affairs. If we visualize time as a 'tree' – where the 'trunk,' the fixed and singular past, splits at the present into a virtually infinite number of 'branches' representing possible futures – we see that the present truth about tomorrow's (potential) sea battle entails that each possible future must reflect that truth; each branch must concur with the present truth. The future thus becomes a sort of restricted domain in which it is a necessary truth (or falsity) that the battle takes place, and so the battle has a kind of limited or conditional necessity that I call 'inevitability.'

And of course, this reasoning applies generally to all future events. For every state of affairs s and every future time t , it is either true or false at present that s obtains at t , and so every future event is inevitable. But surely no event is *necessary*¹⁷ – even if it is actually the case, and so inevitable, that there will be a sea battle tomorrow, there

is a possible world in which there is no sea battle tomorrow; there is nothing in the nature of the sea battle that should render its happening a necessary truth, after all. Despite its apparent contradiction, I think this conjunction – that every event is inevitable but not (absolutely) necessary – stands to reason, because it is not necessary for it to be *true in the first place* that a given future event will occur, in which case the world does not *have* to be as it is. Still, 'as soon as' the world exists, as it were, propositions about the future *are* true. The result is that there can be no partial actualization of the world: once anything is actual, *everything that will ever be actual is already actual*. In this sense, the inevitability of the future is not the inevitability of something 'coming'; the future is inevitable because it is already 'here,' already existent, already *actual*.

It seems to me that inevitability is what the incompatibilist's objection denounces. If T reports that s obtains, then the actual world is a possible world in which s obtains, and so s is inevitable for the simple fact that it is logically impossible, given the previous truth that s would obtain, that s should fail to obtain. But this fact is precisely what the incompatibilist's objection has been raised against – he is complaining that he has no freedom to prevent what actually happens from actually happening. It seems that this is tantamount to complaining that a person lacks the capacity to actualize a state of affairs such that $(A \ \& \ \sim A)$; that is, when the incompatibilist objects in this context that he is unable to act other than as he acts, he is objecting that he is unable to both take a certain action and not take that certain action at the same time. This objection is perplexing at best, for it is a dubious conception of free will which requires that a person be free to actualize contradictory states of affairs. The more appropriate way to understand the inevitability of states of affairs is, I think, to acknowledge that while whatever happens happens inevitably, free-willed beings are free within reason to determine the nature of 'what happens,' to assign content to a state of affairs s . Indeed, this may be a useful way of conceiving of free will: the capacity to define partially the content of states of affairs which are inevitable in themselves yet still receptive to assignment of content.¹⁸

Whatever the case, the heart of the reconciliation of fatalism with free will is simply this: that it is conceivable (and more to the point, it seems to me, conceptually advantageous) that four-dimensional spacetime, as represented by T , is 'closed' and determinate for all times t , but also in some way *closed 'around'* the volitional output of free-willed beings. I present this thesis not as a final and exhaustive report on time and human action, but simply as the preliminary theoretical underpinnings of what I take to be a reasonable account of human existence in a physical universe.

REFERENCES

- Aristotle. (2002). *De Interpretatione*. New York, NY: Oxford University Press.
 Quine, W. (1960). *Word and Object* Cambridge, MA: The MIT Press.

NOTES

-
- ¹ The discerning theorist may hold that these are equivalent means of expressing the same problem, or at least that a solution to one is a solution to the other. Incidentally, for the philosopher of religion, one potential advantage to my approach here is that it relates (if indirectly) to the 'problem' of divine foreknowledge but also demonstrates that the problem is manageable or resolvable even without recourse to the supernatural. Thus if the theologian or philosopher of religion sees no way to reconcile free will and divine foreknowledge via some *religious* or *theological* account, she might be able to reconcile them through the sort of 'secular' account found here.
- ² This principle is operative whether or not the future is thought to 'exist' in a strict sense – that is, whether or not the ontological status of future objects, events, and states of affairs is commensurate with the ontological status of present objects, events, and states of affairs.
- ³ And so the name implies that time is the 'fourth dimension,' not that it is itself four-dimensional.
- ⁴ This view has been variously termed 'block time,' 'block universe,' and 'eternalism.' It is advanced by, for instance, J.J.C. Smart and opposed by presentists such as A. N. Prior. In scientific and mathematical contexts, the four-dimensional theory is sometimes attributed to Hermann Minkowski.
- ⁵ Of course, I write of these things only for illustrative purposes; I am not proposing that there actually *is* a temporal 'line' in the geometric sense, but only that this is a valuable way of conceiving of time.
- ⁶ See Quine (1960, pp. 170-173).
- ⁷ Since a state of affairs – the world at some value of t – is a simultaneity, and since relativity theory implies that a given point in time may correspond to different simultaneities depending on one's frame of reference, the timeline's 'report' may be a function of two input values: time and frame of reference. The timeline fails to report if no frame of reference is specified, and this is not a difficulty for the timeline construct because, according to standard interpretations of Einsteinian physics, any succession of events in time *presumes* some frame of reference in the first place.
- ⁸ At least to the extent to which states of affairs are describable, if not fully.
- ⁹ By 'history' I mean something like the totality of states of affairs throughout the duration of a universe, ordered chronologically. The first part of this formulation seems innocent enough; the notion of 'everything that happens' is theoretically – if not practically – intelligible. But the second part, the chronology, may prove problematic in the wake of relativity theory and its allowance for genuine and irresolvable discrepancies among various frames of reference. Two perspectives order events differently (that is, contradictorily), and there is no method for determining which is the more 'correct' – rather, there *is* no 'more-correctness' to be determined at all. But my concept of a history makes no pretense to objectivity or univity in the sense denied by relativity theory; the conception of a history may be expanded to include not just all actual states of affairs from the 'standard' (geocentric) frame of reference, but from every possible frame of reference as well (see note 7). In this way, rather than espousing some hypothetical 'uniquely correct' chronology, a history can account for discrepancies among equally valid chronologies and still aspire to a consistent and complete description of reality. I believe this solution also

speaks to the (same) problem relativity causes for the *first* part of my definition of a 'history': states of affairs represent simultaneities, and thus are obvious victims of relativity, but expanding the history in the same way – exhaustion of possible frames of reference – affords the same reconciliation.

The history is growing quite complex, but I already hypothesized that a timeline relates a potentially infinite amount of information. Throughout all this, I note that the history is becoming much like what, in the intuition of many, the universe is, at least phenomenologically. (Perhaps this makes sense, and perhaps this theorizing has come full circle, since in everyday speech 'history' often seems to mean something like 'the part of reality which is past' or 'the universe up until now'.)

¹⁰ By this qualifier I mean to exclude possible worlds in which time is nonlinear or nonexistent, if indeed such worlds are possible. The coherence or not of such esoteric forms of time is well beyond the scope of this essay.

¹¹ I will borrow the compatibilist-incompatibilist distinction from the determinism dialogue. The compatibilist believes that human free will is compatible with a fixed future, the incompatibilist contrariwise.

¹² I mean this in a common, everyday sense. This paper is exerting the same sort of power in relation to the reader's cognition (or, by extension, I am). In ordinary circumstances, a speaker seems to partially 'control' or 'determine' the audience's aural experience, etc. I am reluctant to describe this phenomenon in terms of *causation* (hence note 14).

¹³ It is intuitive in that 'it feels like' we have free will, or in that it is 'natural' to believe that we do. In itself this is no conclusive argument for free will, of course.

¹⁴ Where 'primacy' denotes the property of determining the other. If the history has primacy, then agent *p*'s performing action *x* is a sufficient condition for *T*'s reporting that *p* performs *x*. The latter is 'because of' the former.

¹⁵ This formula and the previous are logically equivalent. Elementarily:

1. $\Box(W \rightarrow W)$
2. $\sim\Diamond\sim(W \rightarrow W)$
3. $\sim\Diamond(W \& \sim W)$

¹⁶ Aristotle, *De Interpretatione*, ch. 9

¹⁷ Or, at least, only a few peculiar events are necessary. This matter is beside the point.

¹⁸ I mean this more as an analogy or a metaphor for the method whereby free beings determine what happens inevitably, not as a rigorous account of what it means to have free will. A free agent creates events in history like a logician creates an interpretation for a set of logical formulae.