

# Why I Don't Understand 'Determinism'

## Abstract

I show that the usual definitions of 'determinism' have surprising and previously unrecognized implications concerning the nature of time. I begin by showing that standard definitions strictly imply or strictly preclude an *eternalist* theory of time. This result is perplexing because philosophers who use the term 'determinism' typically say 'determinism' expresses a thesis that doesn't strictly imply or preclude eternalism. I consider possible emendations but show that each falls prey to a similar problem. The upshot is that *either* determinism implies far more than we realized, or debates that make use of the term 'determinism'—such as debates over the compatibility of free will and determinism—have not been adequately defined. This result advances and clarifies inquiries into the nature of free will and the nature of time.

## Why I Don't Understand 'Determinism'

"Some philosophers say they do not know what the thesis of determinism is" – P.F. Strawson

I am one of them. I am not sure what 'determinism' may mean, and I will attempt to explain why. My strategy is to showcase what I take to be the best attempts at defining 'determinism'. I'll show that the best attempts are inconsistent with what philosophers *say* about determinism. The result is that either we have been wrong about what determinism implies, or, if I'm right, we lack a definition that clearly and precisely expresses what the term is supposed to mean. In either case, we learn something significant: the usual definitions of 'determinism', when unpacked, entail unexpected answers to salient questions about the nature of time.

### 1. Defining 'Determinism'

Let us begin with a simple, intuitive definition:

(D<sub>0</sub>) 'Determinism' =<sub>df</sub> 'The thesis that there is exactly *one way* events can go, given the past and the laws of nature'.

This definition may seem to capture what philosophers have in mind—in a rough and ready way. In fact, we'd be done, except the phrase 'one way events can go' is open to interpretation. What does that mean exactly—or roughly? To answer this, consider the following canonical definition, offered by Peter van Inwagen (1983: 85):

(D<sub>1</sub>) 'Determinism' =<sub>df</sub> 'The thesis that

- (i) For every instant of time, there is a proposition that expresses the state of the world at that instant;
- (ii) If  $p$  and  $q$  are any propositions that express the state of the world at some instants, then the conjunction of  $p$  with the laws of nature entail  $q$ .<sup>1</sup>

This definition makes use of the technical term ‘state of the world’. So, how shall we define this term? Not easily, it turns out.<sup>2</sup> I will consider three definitions of ‘state of a world’ and explain why each poses a problem for the above definition of ‘determinism’.

**First Try:** (S<sub>1</sub>) ‘ $s$  is a state of the world’ =<sub>df</sub> ‘ $s$  is a maximal state of affairs’, where a ‘maximal’ state of affairs is a state of affairs that [tenselessly] entails or precludes every other state of affairs.<sup>3</sup>

To see the problem with this definition, suppose, first, that *eternalism* is true, where eternalism is the thesis that there are past and future objects, such that whatever exists at any time (past or future) *exists* simpliciter. Then there is exactly *one* maximal state of affairs that could ever obtain—namely, the one that includes what happens at every time. So, if eternalism is true, then determinism *trivially* follows (according to the above definition of ‘determinism’), since the one state of affairs includes everything about the future. Surely, that can’t be right: determinism isn’t supposed to be a thesis that is *analytically* entailed by eternalism. Philosophers say that eternalism is strictly compatible with indeterminism.<sup>4</sup>

Let us assume, instead, that eternalism is not true. Then the total state of the world is constantly changing. It follows that no state of the world entails any other. To see why, suppose a

---

<sup>1</sup> Cf. Clarke 2003: 4.

<sup>2</sup> Van Inwagen leaves the term ‘state of the world’ undefined but points out conditions that the definition must meet for his purposes (see, for example, 1975: 176).

<sup>3</sup> Cf. Merricks 2007: 127.

<sup>4</sup> Cf. van Inwagen 1975: 176.

state of the world  $\alpha$  entails a state of the world  $\beta$ . Then necessarily, if  $\alpha$  obtains, then  $\beta$  obtains. But since  $\alpha$  and  $\beta$  are both *maximal* (by the definition of ‘state of the world’), it follows that the one obtains if and only if the other does not. Hence, non-eternalism *trivially precludes* determinism. Again, the result can’t be right: determinism is not supposed to be a thesis that is analytically *precluded* by non-eternalism.<sup>5</sup>

Let us try again:

**Second Try:** (S<sub>2</sub>) ‘*s* is a state of the world’ =<sub>df</sub> ‘*s* is a state of affairs that specifies (i) all the objects there are, (ii) what “non-tensed” properties those objects have (such as *being red*, as opposed to *will be red*), and (iii) the relations the objects stand in to one another.’<sup>6</sup>

The thought here is that we should not build facts about what *will* happen into a state of the world. Nevertheless, this definition is no better than the first. For, again, if eternalism is true, then there is exactly *one* state of the world that obtains simpliciter—one which specifies the properties and relations of all objects at all spacetime locations. So, eternalism *trivially entails* determinism, which is probably not a result that eternalists will consider to be the right result of a proper understanding of ‘determinism’. Furthermore, if eternalism is not true, then the state of the world is constantly changing, and thus, once again, no (maximal) state of the world entails any other. In that case, non-eternalism *trivially precludes* determinism, which can’t be right, either.

The third try is a charm:

**Third Try:** (S<sub>3</sub>) ‘*s* is a state of the world’ =<sub>df</sub> ‘*s* is a state of affairs that specifies (i) what objects there are *at a particular instant*, (ii) the non-temporal properties those objects

---

<sup>5</sup> The assumption that presentism may be conjoined with determinism is commonplace. To cite one source, see Bourne 2005: 7–8.

<sup>6</sup> Cf. Rea 2006: 513.

have *at that instant*, and (iii) the relations the objects stand in to one another *at that instant*.

Or: maybe not. There is a problem with interpreting ‘at instant  $t$ ’ on the *presentist* theory that all objects whatsoever *presently exist*? In what sense can an object exist or have properties or stand in relations at a *non-present* time, if all objects exist exclusively at the present time? The only answer I’ve seen that makes sense to me is this one: ‘object  $x$  is  $\Phi$  at instant  $t$ ’ = ‘necessarily, if  $t$  obtains, then  $x$  is  $\Phi$ ’.<sup>7</sup> I think this is the right definition for presentists. But it causes a problem for defining ‘determinism’. To see the problem I have in mind, notice that on this definition of ‘ $x$  is  $\Phi$  at  $t$ ’, if  $x$  is  $\Phi$  at an instant  $t$ , then *necessarily*, if  $t$  obtains, then  $x$  is  $\Phi$ . Therefore, assuming modal system S4, it follows that necessarily, necessarily, if  $t$  obtains, then  $x$  is  $\Phi$ .<sup>8</sup> From here, we infer that necessarily,  $x$  is  $\Phi$  at  $t$  (by replacing “necessarily, if  $t$  obtains, then  $x$  is  $\Phi$ ” with “ $x$  is  $\Phi$  at  $t$ ,” per the definition). So: if  $x$  is  $\Phi$  at an instant  $t$ , then necessarily,  $x$  is  $\Phi$  at that  $t$ . And therefore, if  $x$  is  $\Phi$  at  $t$ , the state of affairs that specifies that  $x$  is  $\Phi$  at  $t$  necessarily obtains. (I am assuming here, for the sake of the proposed definition, that there are states of affairs that specify such things.) It follows that every state of the world *necessarily obtains* (again, because if things are  $\Phi$  at an instant, then that’s necessarily so)—and therefore, every state of the world is trivially entailed by every other state of the world. It follows, then, that presentism trivially entails determinism (on the proposed definitions and assuming S4).<sup>9</sup> Clearly, something has gone wrong.

---

<sup>7</sup> See Merricks 2007: 127; cf. [author’s article]. I don’t claim that this definition will be acceptable to everyone (for example, you might think that the time of your birth could have obtained even if you were not born). But at least its basic terms are intelligible.

<sup>8</sup> In particular, I am assuming the schema,  $\Box P \rightarrow \Box \Box P$ .

<sup>9</sup> There may be a deeper problem here. Presentists who take *times* seriously are committed to there being a sense in which non-present times—or something like times—can exist without obtaining. So, for example, presentists who take times seriously can refer to the Twentieth Century without supposing that the Twentieth

Perhaps these problems can be avoided by making a few strategic adjustments to the van Inwagen's definition. We want to be able to talk of states of the world without such states *trivially* entailing or precluding all others. So, let's try this:

(D<sub>2</sub>) 'Determinism' =<sub>df</sub> 'The thesis that

- (i) For every instant of time, there is a proposition that expresses the state of the world at that instant;
- (ii) If *p* and *q* are any propositions that express the state of the world at some **distinct** instants, **t<sub>1</sub> and t<sub>2</sub>, respectively, where t<sub>1</sub> is earlier than t<sub>2</sub>**, then the conjunction of *p* with the laws of nature entail *q*, where
- (iii) 'x is a state of the world' = 'x is [or is expressible by] a proposition that specifies (i) all the objects there are, (ii) what non-temporal properties those objects have, and (iii) the relations the objects stand in to one another.

This definition combines the second attempt at defining 'state of the world' with a few bolded modifications to the second clause of van Inwagen's definition. The purpose of the modifications is to restrict the scope to states of the world that don't *trivially* preclude or entail all other states of the world.

But there is still the same basic problem: if eternalism is true, then there is exactly *one* state of the world that obtains (one which specifies the properties and relations of all objects at all spacetime locations), and therefore, no states of the world can occur *after* any other. So, this time, eternalism *trivially precludes* determinism. That can't be right, either.

One more try:

---

Century *obtains*. The same should be said of instants: instants are duration-less parts of times, and non-present instances of the Twentieth Century exist but no longer obtain. But then instances are functionally equivalent to *states of the world*. And then the definition of 'state of the world' in terms of 'instance' appears to be circular.

(D<sub>3</sub>) ‘Determinism’ =<sub>df</sub> ‘The thesis that *if* eternalism is true, then the conditions given by (D<sub>1</sub>) & (S<sub>2</sub>) are met; otherwise, the conditions given by (D<sub>2</sub>) are met’.

This definition neatly solves all the problems we have seen; and I have not been able to come up with any other definition that solves those problems.

However, we shouldn’t take (D<sub>3</sub>) seriously for a simple reason: ‘determinism’ isn’t a theory about the implications of certain *theories of time*, whereas (D<sub>3</sub>) is.<sup>10</sup> Moreover, our understanding of ‘determinism’ is not supposed to depend upon a prior understanding of ‘eternalism’. Yet, we cannot understand (D<sub>3</sub>) without understanding ‘eternalism’.

The problems we are facing are deeply rooted and afflict the current definitions of ‘determinism’ on offer. Take, for example, David Lewis’ definition (1983: 360) in terms of ‘divergent worlds’: a world *w* is deterministic *iff* no worlds that diverge from *w* conform perfectly to *w*’s laws. The problem here is with unpacking what it means for worlds to diverge. Presumably, divergent worlds share a state of the world at one time and fail to share a state at a later time. But this account simply takes us back to van Inwagen’s definition, and the same problems arise: if ‘state of the world’ is defined in a time-indexed way, then it turns out that presentism trivially entails determinism (recall (D<sub>2</sub>)), and if ‘state of the world’ is not time-indexed, then eternalism trivially entails determinism. The definition only *seems* to work if we mentally hold fixed whether eternalism or presentism is true. Yet if the debate over these theories of time is indeed orthogonal to the debate over determinism, as philosophers suppose, then Lewis’ definition fails. All other definitions I’ve encountered fail for this same reason.

---

<sup>10</sup> This definition also has the ironic implication that *Peter van Inwagen* doesn’t understand determinism, since van Inwagen says he doesn’t understand *eternalism* (because he doesn’t grasp, or claims not to grasp, any meaning of ‘is’ that isn’t *present tense*).

Maybe someone will be able to find a successful definition that avoids the above problems. That would be good news. But the current news is that such a definition has yet to be given. So we face a challenge: develop a decent definition of ‘determinism’.

## 2. A Concession

Perhaps I have overstated the results. Perhaps what I have shown is *not* that philosophers fail to grasp the meaning of ‘determinism’. Rather, what I have shown, more exactly, is only that certain technically precise definitions have surprising metaphysical consequences. The consequences are surprising because philosophers often *say* that determinism doesn’t have those consequences. But maybe the problem is with our understanding of the *implications* of determinism rather than with our understanding of the *meaning* of the term, ‘determinism’.

I concede that I might be going too far to suppose that we don’t understand ‘determinism’. My reason for suggesting that we fail to see clearly into the meaning of ‘determinism’, rather than that we fail to see clearly into the implications of *determinism*, is that it seems to me that if we grasped the meaning of ‘determinism’, then we wouldn’t have said things about determinism that are logically inconsistent. I’m operating out of a principle of charity: if smart people say something that, to my ears, is logically inconsistent, I probably don’t know what they mean. But I don’t expect everyone to go along with me here. It could be that philosophers, some of them anyway, do indeed grasp the meaning of ‘determinism’ well enough—or that D0 is intelligible in its own right—without a precise unpacking. (Compare: we grasp the meaning of ‘chair’ well enough, whether or not we can give an adequate definition of ‘chair’.) So, I shall concede here that as far as I have argued, there *may* be philosophers who understand the meaning of ‘determinism’.

Even still, my investigation yields a significant conclusion. Suppose there is a determinable (or pre-philosophical) meaning of ‘determinism’ that philosophers have in mind when they use the term ‘determinism’. We have seen that specific *determinates* of the theory expressed by ‘determinism’ have startling implications about the nature of time.<sup>11</sup> In particular, certain precise interpretations of van Inwagen’s canonical definition yield a theory that is trivially entailed by eternalism. Other interpretations yield a theory that is trivially entailed by presentism. Furthermore, I found no precise interpretation of any definition that is neutral with respect to both presentism and eternalism. This result is strange and startling. It implies that our most *precise* interpretations of ‘determinism’ have substantial implications about the nature of time. This news should be of great interest to philosophers of time.

### 3. Implications

I have tried to show that the canonical definition of ‘determinism’ contains key terms that cannot be properly defined in a metaphysics-neutral way. I have done my best to come up with a good definition myself—one that doesn’t trivially imply or preclude substantive philosophical theories about time. But I have come up short.

The implications are grave. Consider that ‘determinism’ is a central term in a number of philosophical conversations about free will. Yet the meaning of the term is not pre-philosophically obvious, like ‘up’, ‘happiness’ or ‘true’. So, if we are to clearly grasp what it means, it seems we should have at least some idea how to explicitly define the term. But we don’t—or at least, it seems no *precise* definition is metaphysically neutral in the way that philosophers have thought. I suggest two possible lessons. One lesson—one I am sympathetic

---

<sup>11</sup> Dieckemper 2007 displays a similar dialectic when considering how to unpack the meaning of ‘fixity of the past’ in a metaphysics-neutral way.

with—is that much of the “free-will” literature is built on straw: all the arguments for and against the thesis that free will is compatible with determinism, for instance, fail to rise to the level of being intelligible, since ‘determinism’ fails to be intelligible. (That isn’t to say that everything that goes by the name “determinism” is unintelligible. I leave open whether there may be intelligible *causal* or *ontological dependence* notions of “determinism”. In fact, one possible take-home lesson is that we ought to cast all the compatibilist debates in terms of causation or grounding. Cf. Todd 2012.)

Alternatively, the lesson is that although some of us may grasp, to some extent, the meaning of ‘determinism’, there are actually different specific versions of determinism corresponding to different specific definitions of the term. In that case, the debates over compatibilism can now be recast in terms of the different versions of determinism. So, for example, someone might think that free will is compatible with the version of determinism that is trivially entailed by eternalism (Try 1), while thinking that free will is *incompatible* with the version of determinism that is trivially entailed by presentism (Try 3). Clearly, the debates over free will would then become inextricably entangled with debates over the nature of time.

I would like to close by offering a third diagnosis for consideration. Recall that our problem has to do with finding a definition of ‘determinism’ that works on *both* eternalist and non-eternalist theories of time. Maybe the lesson is that eternalists and non-eternalists do not actually mean the same thing by ‘determinism’. Here is evidence. For many of us, the first time we heard about eternalism, our *instinct* was to think that eternalism implies determinism; we imagined that all events, including future ones, would be *laid out* from time eternal. And we thought, “in that case, my future would already be *set*—thus, determined!” But of course, we became wiser: we came to recognize a distinction between time and necessity, and we came to

understand that (non-causal) indeterminism is a thesis about temporal recombination: past states plus laws don't *necessitate* future states. In other words, indeterminism is compatible with eternalism, we thought. But maybe we thought wrong. Maybe our initial, naïve instinct was closer to the mark—from *the presentist perspective*. Think about it this way. When the presentist says, “past states necessitate future states,” she must *mean* something different than what the eternalist means by “past states necessitate future states”. For if they meant the same thing, then there would be an adequate definition of ‘determinism’ agreeable to both parties. But there isn't.

## References

- Bourne, C. 2005 "A Theory of Presentism," *Canadian Journal of Philosophy* 36: 1–23.
- Clarke, R. 2003. *Libertarian Accounts of Free Will*. New York: Oxford University Press.
- Diekemper, J. 2007. "B-Theory, Fixity, and Fatalism," *Nous* 41: 429–52
- Lewis, D. 1983. "New Work for a Theory of Universals," *Australasian Journal of Philosophy* 61: 347–377.
- Merricks, T. 2007. *Truth and Ontology*. Oxford: Oxford University Press.
- Rea, M. 2006. "Presentism and Fatalism," *Australasian Journal of Philosophy* 84: 511–24.
- Strawson, P. F. 1962. "Freedom and Resentment," *Proceedings of the British Academy* 48: 1–25.
- Todd, P. 2012. "Soft Facts and Ontological Dependence," *Philosophical Studies Online* First: 1–16.
- van Inwagen, P. 1975. "The Incompatibility of Free Will and Determinism," *Philosophical Studies* 27: 185–99
- \_\_\_\_\_. 1983. *An Essay on Free Will*. Oxford: Clarendon Press.