LSE CPNSS

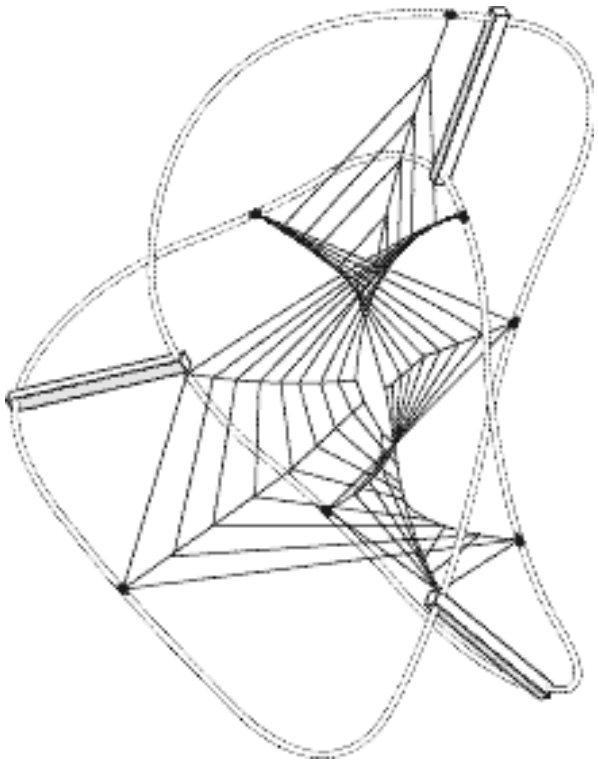
Centre for Philosophy of Natural and Social Science

Measurement in Physics and Economics

Technical Report 16/01

Freedom From the Inside Out

Carl Hoefer



Editor: Peter Dietsch

Freedom from the Inside Out

Carl Hoefer, LSE
Draft 2, March 2001

0. Introduction

Since the death of strong reductionism, philosophers of science have expanded the horizons of their understandings of the physical, mental, and social worlds, and the complex relations among them. To give one interesting example, John Dupré has endorsed a notion of *downward causation*: "higher-level" events causing events at a "lower" ontological level. For example, my intention to type the letter 't' causes the particular motions experienced by all the atoms in my left forefinger as I type it. The proper explanation of the motions of an atom at the tip of my forefinger primarily involves my intentions, rather than (for example) the immediately preceding motions of other nearby atoms, or any other such particle-level events.

While this is a natural enough idea on the face of it, such downward causation has seemed to be in tension, or outright conflict, with another compelling intuition, which Dupré calls *causal completeness*.

> "This is the assumption that for every event there is a complete causal story to account for its occurrence. Obviously enough, this is a view of causality the roots of which are to be found in the soil of determinism. The paradigm of a complete causal story is the sufficient (and perhaps even necessary) antecedent condition provided by a deterministic causal explanation. However… [since microphysics seems likely to be indeterministic], it is important to consider the indeterministic analogue of deterministic causal completeness. It is not hard to see what this should be. The basic idea is that there should be some set of antecedent conditions that together determine some precise probability of the event in question." (Dupré 1993, pp. 99 - 100)

As Dupré points out, belief in some such doctrine as causal completeness underlies the attraction many philosophers still feel for reductionism, despite the latter's untenability in any form stronger than mere supervenience.

It is causal completeness that is at the heart of the age-old dichotomy between free agency and physical determinism. For if determinism is true, there is a prior, sufficient cause of my fingertip's atoms' motions: the earlier state of the physical world. "Earlier" could mean mere moments ago, or it could mean at some time before I was born. Given this prior, sufficient cause, my intentions seem idle and epiphenomenal; they are there, to be sure, but they are *just as much* caused by this prior physical state of affairs; my "free" will seems then a hollow joke.

But determinism is not necessary for the threat of causal completeness to free agency

to arise. For reasons that Kant first realized, indeterminism at the microphysical level does not seem to help. The randomness, if any, in microscopic phenomena does not seem to "make room" for free will, but rather only replaces a sufficient physical cause with (at least in part) blind chance. The presumption in favor of *upward* causation and explanation (from microphysical to macrophysical) that comes with causal completeness is what cuts free agency out of the picture, whether this causation is deterministic or partly random.

Philosophical subtlety has thus put our freedom in double jeopardy.[1] Some philosophers of a pluralist/empiricist bent (Cartwright, Dupré for example) respond by saying that they trust the evidence of common sense more than such philosophical subtleties. Maybe there are really no laws of nature at all, in the strict sense; maybe causation at the micro-physical level deserves no priority over causation in the form of human agency. At any rate, on any viable concept of *evidence*, they say, the evidence in favor of free will is stronger than that for universally true physical laws.[2]

On this last point, I think they are wrong. We have overwhelming evidence for universal, exceptionless laws of nature, if we don't stack the deck (i.e., the requirements of evidence) against laws in the way Cartwright does. But fortunately, skepticism about true, universal laws of nature is not necessary to derail the apparent challenge to free agency coming from causal completeness. All that is needed is a proper understanding of *time* – what it is in the physical world, what it is in human affairs, and how they are related. Given the proper understanding of time, we will see that freedom and determinism are compatible – compatible in a much more robust sense than has ever been thought possible.[3]

1. The Two Times

"Time" means one thing in physics, and something quite different in everyday human affairs. McTaggart first described the distinction clearly, and gave the two times names: A-series time and B-series time.

A-series time is the time in which we live our lives. There is the *past*, the *present* (the "now"), and the *future*; and the present "moves" inexorably into the future leaving more and more of our lives behind us. B-series time is by contrast "static": time is a linear ordering

[1]There is of course a long tradition of philosophers responding to this apparent threat by arguing for the compatibility of freedom and determinism, when the former is properly understood. For what follows, I need not enter into these debates. See Fischer (1994) for a comprehensive and robust defense of (a form of) compatibilism.

[2]Dupré does in fact say exactly this, in his paper "XXXX". Cartwright has not discussed free will explicitly, to my knowledge, but her views about laws, causation and evidence seem to fit well with this response.

[3]On a topic such as freedom of the will, it is too much to hope that any proposal can be completely novel. In section 4 I link my proposal to Kant's reconciliation of freedom and determinism. Further, as I recently discovered, Peter Forrest (1985) has defended free will along lines similar in many respects to those developed here. There are however quite substantial differences, and in particular Forrest does not bring considerations about *time* into play in his account.

(or partial ordering), typically represented by a line on which each point represents an instant of time, but no point is distinguished as "now" (see figure 1). *Things* may change in B-series time, by having one set of properties at one point, and a different set of properties at a later point. But time itself does not "change" or "move". Physics seems to describe the world entirely in B-series terms, and to have no need of A-series concepts such as *present* and *future*. Indeed, many philosophers believe that physics since Einstein's 1905 relativity theory is outright *incompatible* with the A-series.[4] Regardless of whether this is correct or not, it is still true that physics does not *require* A-series time notions, and seems to find a natural fit only with B-series time. (A possible exception to this is in quantum mechanics, but only under the most bizarre (many-world) or idealistic (consciousness-collapse) interpretations.)

When space is combined with B-series time explicitly, as Minkowski first did in 1908, we get a description of the world as a whole, with four dimensions. This is what we do in drawing "Minkowski diagrams" in relativity theory, but it works equally well from the perspective of Newtonian physics. Either way, philosophers have found it useful to think of the world, consisting of 3 spatial dimensions and one (B-series) temporal dimension, as a "block universe" (see figure 2).

In the block universe, time is certainly to be singled out as *different* from the three spatial dimensions – in terms of the laws (if any), the metrical structure(s) giving spatial and temporal distances between events or world-points, and so on. Likewise, at least over the part of the block accessible to our observations, the two *directions* of time (past-directed / future-directed) are distinguishable. But what is not to be found is an ontological separation of parts of the block into past, present and future. This striking fact means that events of 1000 years in Earth's future are, in terms of reality or existence, no different from the events (*now*) of your reading these words, or the events of last week.

This notion is hard to grasp, and feels threatening to us as free agents. It has even been advanced, incorrectly, as a vindication of fatalism. For, viewing ourselves and our actions from within the A-series perspective, we think of future events as *open* in some real sense, to be determined (partly) by *our* choices. But in the block, all events are equally real, those in your far future no different from those in your past.[5] This 4-D block world that physics offers us *seems* impossible to reconcile with this agent-centred, A-series-embedded perspective.

I will now argue that in fact, matters are just the opposite of how they seem. The very "timelessness" of the 4-D block (in an A-series sense) leaves us free to *reject* the customary

[4]The reason is that the A-series seems to require a privileged way of dividing events into those happening *now*, vs. those in the past or future, which is effectively a privileged notion of absolute simultaneity. Special relativity, as standardly interpreted, is incompatible with an absolute standard of simultaneity.

[5]See Horwich (1987), *Asymmetries in Time*, for the correct refutation of the argument for fatalism ("logical" fatalism) based on the block universe.

view that *past* events determine present choices. From the B-series perspective there is no reason to think of past ⇙ future determination as more important or real than future ⇙ past determination. And, even more to the point, one can equally view a set of events in the *middle* as determiners of both past and future events.

This is exactly what we should do. Our *free* actions, intentions, thoughts etc., in the middle of the block universe, are *part* of what determines how the rest of the block shall be. In order to make the point as clearly as possible, I will first discuss things under the assumption of some standard, Newtonian-style determinism. The idea here is that given the complete state of affairs "at a time" in the universe (i.e., all physical facts specified on a time slice or thin sandwich), plus the true laws of nature, all earlier and later physical events are logically determined.[6] Weaker forms of determinism can be defined, but they pose, *prima facie*, less of a problem for free agency.[7] Later I will come back to freedom in a causally complete but probabilistic world.

2. Freedom from the Inside Out

Determinism tells you that the state of the world at a time determines all the rest, past and future, but it doesn't tell you *which* slice, if any, explains or determines all the rest. The challenge to free will from determinism has not come from the physics, but rather from the unholy marriage of deterministic physics with our A-series view of time. The worry we have is that a *past* slice (long in the past, maybe even the "initial conditions" of the universe if there are such) determines our actions *now*. We never think of a now-slice (including the voluntary actions we perform now) determining what happened in the past. Why not? There are two reasons. First, we unconsciously assume a metaphysical picture that is A-series based and incompatible with the block universe: we think of the past as "real", fixed or determinate, the present as also "real" (or becoming so), but the future as "indeterminate" or "open". And as the zipper of the now moves into the future, it's the future that is getting determined, not the past. Once one unearths this metaphysical lurking picture, its irrelevance becomes obvious. Physics has no truck with any of it, and (as noted before) is probably incompatible with it, when understood as applying to physical events *per se*. From the B-series or block-universe perspective, there is no reason to think of the past as determining the present and future, rather than vice-versa, and so on.

---

[6]See Lewis (1994) for an explication of determinism in terms of possible worlds, and Earman (1986) for detailed discussion of the difficulties of defining and assessing determinism in various physical theories. The strong "Newtonian-style" determinism I am assuming for the moment turns out, as Earman shows, to be best captured not in Newtonian physics but rather in Special-relativistic physics. For the discussion to follow, I am assuming bi-directionality of determinism. This is assured by time-reversal invariant physical laws, but this condition is not needed. For example, Callender (2000) argues that QM under a Bohmian interpretation is not time-reversal invariant. It is nevertheless bi-directionally deterministic.

[7]See section 2.1 below for further discussion of this point.

The second reason is more interesting. When we consider the idea of events in a time slice *now* physically determining the past, we become nervous because it looks as though we are positing *backward causation*. So if I assert that my actions, now, are free and explained or determined only by my own will; and that in a deterministic world, that may entail consequences about the past, but so what?; it looks as though I am positing backward causation, and giving myself the power to affect the past. And this is thought to be unacceptable on solid *physical* grounds, independent of any A-series/B-series considerations. In many presentations of the incompatibility of determinism and free will, this worry about affecting the past comes out explicitly: saying "I could have done otherwise" is analyzed as tantamount to saying "I could either have caused a law of nature to be violated, or changed the past."[8] Laying to rest this worry that freedom with determinism must involve either unacceptable backward causation or "changing the past" will be the first task of the next section.

The idea of freedom from the inside out is this: we are perfectly justified in viewing our own actions *not* as determined by the past, *nor* as determined by the future, but rather as simply determined (to the extent that this word sensibly applies) *by ourselves, by our own wills*. In other words, they need not be viewed as *caused* or *explained* by the physical states of other, vast regions of the block universe. Instead, we can view our own actions, *qua* physical events, as primary explainers, determining – in a very partial way – physical events outside ourselves to the past and future of our actions, in the block. We adopt the perspective that the determination or explanation that matters is from the *inside* (of the block universe, where we live) *outward*, rather than from the *outside* (e.g. the state of things on a time slice 1 billion years ago) *in*. We are free to adopt this perspective because, quite simply, physics – including our postulated, perfected deterministic physics – is perfectly compatible with it.

As I said before, exploring the consequences of this perspective and defending it against apparent problems will occupy most of the rest of the paper. But the key to the defense has already been explained, and needs repeating. The notion of *past* events determining and explaining *future* events, and the opposite direction (or an "inside-out" direction) of explanation being somehow wrong or suspect, arises completely from an unholy marriage of A-series time with deterministic physics. The mistake is natural and understandable, because of the way the A-series dominates our lives and our thinking, especially causal/explanatory thinking. It remains nevertheless a mistake. A deterministic physics gives us *logical* relations of determination, not a unique *temporal* relation of determination. In the block universe one can view a slice now, or a future slice, or a future ½-block, or the past ½-block "before" now, as logically determining the rest.[9] These logical

[8]For example, see van Inwagen (1975). For a thorough and illuminating treatment of the challenge to free will from determinism, see Fischer (1998).

[9]Always assuming the truth of the deterministic laws, of course.

relations however are not in any interesting sense *explanatory*, nor even *causal*. Physics does not pick any one out as more important than the others, and indeed, equally allows us to ignore all of them when it comes to thinking about causation and explanation in things that matter in our lives.

This is not the way we are accustomed to thinking of determinism. We usually stay in our A-series perspective on the world, tacitly conflate determination with causal explanation[10], and there we are, mired in the apparent incompatibility of determinism with our actions' being explained by our choices. A first antidote to this mistake is firmly to keep in mind that physical determinism belongs in the B-series world of physics alone. To break the conflation between determination and causal explanation, it helps to remember that deterministic physics equally allows future ⇦ past determination, but it does not thereby tell us that the future *causally* explains the past. The full antidote can only come by exploring the consequences of an "inside-out" perspective on determination, and making sure that they are acceptable both physically and for common sense.

2.1 Past ⇦ future determinism only? Above I said that weaker forms of determinism than the full time-symmetric Newtonian type we have been assuming pose, *prima facie*, less of a threat to free will. But do they really? In particular, does my argument for freedom from the inside out still have plausibility, if what physics gives us is past ⇦ future determinism but *not* future (or middle) ⇦ past determinism? At first blush, it might seem that the plausibility of the perspective on offer is undermined. A closer look remedies this misapprehension.

"Past ⇦ future determinism only" means that the future ⇦ past relationships allowed by the laws are one-many, while the past ⇦ future relationship is one-one. These relationships are still, however, *logical* rather than causal or explanatory. As long as our physics remains fully expressible in terms of B-series time, and has no need of A-series time, the one-way character of their determinism does *not* mean that the past is "fixed" in some sense *vis-à-vis* future events. Nor do past events become somehow "logically prior" to present and future events in the block. It is true that we can say that the past (plus the laws) entails our present actions, and can not any longer make the same claim regarding the future (which claim, psychologically, perhaps helped break the grip of the idea that these determination relationships render us unfree.) But this change does *nothing* to weaken the claim that the physical world's time is B-series time, in which past, present and future events all have the same ontological status. It does *nothing* to re-assert the notion that the past is "fixed, done, and beyond our control". In short, because this hypothesized weaker form of determinism does not re-impose an A-series metaphysics of time on us, it does not at all undermine the perspective of freedom from the inside out.

[10]Often the conflation is explicit, as the phrase "causal determinism" indicates.

In fact, in terms of the worries for this perspective that we are about to explore, past ⇙ future only determinism reduces their strength. We are about to consider worries that arise if we consider our free actions as prime, explanatory starting-points, having consequences toward both past and future. But as just noted, under past ⇙ future only determinism, the present - past relationship is one-many rather than one-one. So whatever the constraints our free actions place on the past turn out to be, in principle they will be *weaker* than they would under full, bi-directional determinism. The comment made at the end of §1 was correct: the challenge to free will posed by weaker forms of determinism *is* in fact weaker.

3. Causation and Consequences

Can this "inside-out" perspective be held, though? Does it not make the mistake of claiming that our actions now have causal consequences toward the past? The answer is no. From the inside-out perspective, our freely chosen actions place *constraints* on what the past and future can be like, but the constraints are astonishingly weak, both toward the future and (especially) toward the past.

To discuss the question in more detail, let's assume that a human action (including the perceived surroundings of the agent's context) is a physical event *type* that has innumerable instantiations at the microphysical level. We assume, in other words, that there is some ill-defined and probably infinite set of microphysical state-types that are "good enough" to count as a supervenience base for my typing "t" in the assumed context. In doing so, we are doubtless assuming a more reductionist picture than is likely to be true, but this is needed in order to make the apparent challenge as strong as possible.[11]

3.1 Consequences toward the past

If I freely choose to type this letter, "t", the choice in its context entails that some one of this enormous number of micro state-types shall be, and that is all. The constraints this places on how the past should be, even (say) the past of only one minute earlier, are probably either trivial or non-existent. Thinking of the constraints toward the future helps illustrate the weakness in either direction.

In his famous 1908 paper on causation, Russell pointed out that no cause ever *guarantees* the following of the customary effect – unless we inflate what we count as the cause to make it identical to a time-slice of the whole state of the world over a huge region of space. I reach for the "t" key, I depress it; will a "t" appear on the screen microseconds later? Typically, of course, yes; but not as a matter of logic (plus the laws), unless we rule out all

[11]Here I am giving the benefit of the doubt to supervenience on the microphysical, and interpreting it relatively strongly. The idea is that, although there can be no conceptual reduction of "Carl types the letter "t" on his laptop" to the language of microphysics, it is nevertheless the case that at least God could say, if you showed him a microstate, whether or not it is good enough to count as "Carl typing the letter "t" on his laptop" (or Ct for short), in the context.

possible interferences that could intervene and prevent the effect. (Think of every way that the computer might malfunction, or the power be cut, or a black hole whiz through your CPU at exactly the right time, …)

The same goes toward the past: in terms of logical determination, our actions have little or no necessary consequences about what the past shall be like, outside of what is already presupposed in describing the context. At the microphysical level the constraint is just that earlier microphysical states have to be logically consistent with a microstate of the correct type (i.e., one corresponding to my typing a "t") obtaining, at the time and place that it does. If the microstates we are positing cover, for example, a spatial area of 10 metres radius, then any given microstate logically entails the earlier microstates (i.e., toward the past) over an ever-shrinking spatial region, which vanishes "after" a time period exactly equal to the time that light takes to travel 10 metres (see figure 4). Specifying the microstate over a region of space and a slice or sandwich of time, in other words, logically determines the past and future microstates only over symmetric past- and future-pointing "light cones" which exist only for an absurdly short period of time. All this is so, assuming Special Relativity's restriction on the velocity of physical things. If we remove that restriction, then the regions of past and future logical determination vanish entirely. And when we recall the huge (probably infinite) number of microstates that can serve as basis for a macroscopic event (my typing the "t"), the logical determination toward the past is correspondingly decreased. (When I freely choose to type the "t", I do not thereby choose to actualize a *particular* microstate!)

Despite the correctness of all this in logical/physical terms, we nevertheless have to acknowledge that the region of "practical determination" of the state of things, toward the future, is usually much greater. There are pervasive and fortunate circumstances in the physical world that allow it to be the case that interferences such as Russell noted are rare, and that we are usually successful in producing the effects that we want toward the future. After all, usually my computer is functioning perfectly, there are no black holes or meteors or laser bolts heading toward me, etc. So usually, when I type "t", a "t" appears and stays there for a while. We are able, fortunately, to make things be the way we wish at the macro-level, more or less completely – depending on what we're aiming for – and for a goodly amount of time. If the same thing were true toward the past, then if freedom from the inside out were the case, we should have the ability to freely choose to make past events be the way we wish (most of the time, to some limited extent). This would quickly lead to paradoxes of the time-travel variety. For example, having observed the word "example" on my screen for the past minute, I could (it seems) now take some action that causes the screen to be blank for the *past* minute. This means either postulating a "changing of the past" – which is incoherent, or at the very least takes us outside of the block-universe perspective we have been assuming – or mysterious interventions that prevent us from succeeding in our backward-effect actions.

(Banana peels are the standard mechanism, in the time travel literature.) But fortunately, the same thing (ability to cause large-scale, enduring effects most of the time) is *not* true toward the past.

Temporally asymmetric features of our world make it very unlikely that our free actions leave "traces" on the past of a macro-level and repeatable nature. Philosophers interested in the "direction of time" problem have documented some of these circumstances in depth: the thermodynamic asymmetry, the "fork" asymmetry, the knowledge asymmetry, the radiative asymmetry, and so on. Here is an example. We know that if we want to have a drink be at a uniform 2☜C, we can start with our mixture at room temperature, add lots of ice, and wait. We don't have to worry about the drink getting hotter and the ice bigger. But suppose that on alternate days, the Second Law of thermodynamics switched temporal directions. Then on alternate days, we could cause a nice cool drink to *have been present* earlier, by adding ice to a room-temperature mixture.[12] But things being how they are, thermodynamically, we can't do anything of the sort. I can add ice to the mix, but nothing at all is then entailed about the past features of the drink – even assuming the absence of external influences. It might have sat there for a day, at equilibrium; or it might have had ice in it two hours ago; or it might have been quite hot; and so on.

It appears that our inability to produce causal effects toward the past is largely due to (1) these pervasive asymmetries in physical phenomena; and (2) the nature of our conscious experience and of sensation, which are either "in", or somehow produce, the A-series, flowing time of common sense.[13] I regard these as very puzzling and unresolved issues; fortunately they do not need to be resolved for our purposes here. What matters is that our free actions, while they may have logical *consequences* about the past because of determinism (of a highly disjunctive nature, and for a trivial amount of time), do not have to be thought of as causally *bringing about* large-scale features of the past, or as explaining them.

Finally, notice that if all this was incorrect, the traditional picture would be in trouble also. Suppose we decided that, on the assumption of freedom from the inside out, we *would* in fact be able to effect noticeable backward causation. In that case, merely rejecting the perspective of freedom from the inside out would not automatically make this backward causation go away! Sticking with the same deterministic physics, the physically possible worlds with the backward-causation events would *still* be physically possible even if we stuck to thinking of determinism as a past ⇖ future relation. The human actions producing backwards effects would still be *physically* possible. So if backward causation is a worry for

[12]Here I am glossing over all the thorny problems about whether human bodies could live under such a reversal of thermodynamic asymmetry, and whether the perceived flow of time would not then reverse as well.

[13]See Horwich (1983) and Price (1996) for extended discussion of temporal asymmetries.

one perspective, it should be a worry for both. Further constraints on physically possible worlds would have to be added to eliminate the threat, and their justification would not be from physics alone. But let's leave this concern for now, and carry on assuming that backward causation is not a worry.

3.2 Harmony

Granting this, one still might have some worries about *harmonisation* – about whether all the different actions we believe ourselves to be able to freely choose, can really fit together (a) with each other and (b) with the past as we know it (i.e., the macroscopically described, known past) under determinism. One wonders whether billions of humans, all exercising free will, over the course of millennia, shouldn't be expected to generate enough consequences toward the past to generate contradictions – despite the weakness and disjunctive nature of the consequences of each act taken on its own. My freely chosen actions don't just have to harmonise with my immediate past; they have to harmonise with your immediate past and everyone else's, and they all have to be able to be fit together into a consistent past history of the world. The worry here is this: how do we know that there is always *at least one* microstate of the whole past that is compatible with the consequences (toward the past) of all the freely willed actions of all agents in history? Might it not be, instead, that once the free choices of (say) 4 billion humans are conjoined, then the possible choices of the rest of humanity are either removed (only one overall microstate is compatible) or severely constrained (each of us has few genuine choices available to us)? Given that everything has to fit together in such a way as to not violate the physical laws, one may worry that there needs to be a pre-established harmony (or, better: a harmony *simpliciter* – the "pre" is misleading), and that because of this, we really are not free to do all sorts of things after all.

We know, of course, that all *actual* choices in fact fit together harmoniously; this is our starting assumption, that a deterministic microphysics holds sway over all actual events. So this harmony worry really has to do not just with actual choices, but the *alternative* choices we think we could have made: our freedom to do otherwise. This then brings us to the heart of the issue of the compatibility of freedom and determinism: the counterfactuals we believe, the could-have-done-otherwise's.

When I type the letter "s" I may think that I could have chosen to type a "z" instead, in keeping with my nationality. And I think I could have done so, *with the past being, macroscopically, just the way I know it to be*. But can I really? Or is it instead the case (though we can't of course see why) that for me to type that "z" instead, the past would have to have been different *macroscopically* (e.g., I would have had to have had corn flakes for breakfast instead of toast)?

The qualifier "macroscopically" is absolutely crucial here. For note that although we are sticking to a block-universe perspective when it comes to real physics, and hence not

supposing that the actual physical state of the world this past morning is somehow ontologically privileged over present or future states, nevertheless in terms of *our actions as we conceive them*, there *is* an important asymmetry. We think of ourselves as beings with a certain history, in a physical world with its own history, and our actions as arising freely *given* (or despite) all that.[14] And if this perspective was not in fact sustainable, then the compatibility of freedom with determinism I am after would not be possible after all.[15] I think I have freedom of the following kind: even given that the past history of the world is, macroscopically, as I (and indeed every other agent) knows it to be, I can either type the "s" or the "z" (depending on which I choose). Can the past and our present actions, *as well as those we don't choose but think we could,* all fit together harmoniously in the way this conception of freedom demands? Can everything harmonize as well as harmonise?

Part of the response to this worry is what has already been explained: that logically each person's free actions entail only (at most) that one of an enormous set of past microstates obtain, and that only over a time-span that is vanishingly small. The time-asymmetry of typical physical events further rules out that there should be macro-scale consequences toward the past under "typical" circumstances. If each person's free actions entails practically nothing about the past, it is plausible that all persons' actions conjoined should be able to fit together consistently. Moreover, of course, looking at the actual world from the block-universe perspective, all human actions *do* fit together consistently. So we have one example of a universe where it all works. The worry of course is that *only* one such, or very few such worlds are possible given the laws and the contextual/historical circumstances of our free choices. But what reason can we have for this worry?

If anything, it seems that evidence points strongly the other way. I can test my free will right now, in the very typical circumstance of a person typing on a computer in a small room. I type various letters, randomly. Think of each letter struck as a run of an experiment. The experiment is simply to see whether all sorts of letter-producing choices, in a very normal physical context, starting from macroscopically near-identical initial conditions, can fit together consistently into one history. And the result is clear: they can.[16] Extending this

[14]Here I am implicitly offering a criticism of one standard way (Lewis') of analyzing counterfactual statements. Lewis, who seems inexplicably wedded to the A-series in all his metaphysics, supposes that in most uses of counterfactuals we mean to hold the past fixed – and I agree. But for Lewis this means the *physical* past in all its gory microphysical detail; so if determinism is true, it takes a miracle to get the if-had-done-otherwise scenario started. But why hold the past fixed in microphysical detail? What matters for action is the macroscopic past, that we know about empirically. When only that is fixed, I suggest, we don't need miracles to postulate various different actions and their likely future consequences.

[15]In this case, we would have to live with the threat to freedom posed by causal completeness, or take up a different compatibilist picture, such as that of Fischer (1994). He argues that the freedom-relevant sense of control over one's actions is "guidance control", which does not require the ability to have done otherwise.

[16]The point in this thought experiment is to keep the macroscopically-described initial conditions as much identical as possible: not only the room, lighting, etc. are the same, but also my intention – namely, to type a letter at random.

idea further, we can regard much of what happens in an everyday life as providing similar evidence for harmony between a given, fixed macroscopically-described past and multiple present choices. I go to the 4th-floor cafeteria every day, and the menu on offer is always the same; but my choices vary.

Someone gripped by the harmony worry here will say that all this shows nothing. For each letter typed and each lunch selected may *not* be in fact freely chosen, but rather determined by the requirement of there being a globally consistent history (even when only some of the past, macroscopically described, is held as fixed). I find this worry very implausible, verging on the paranoid. The idea is that somehow, the deterministic physics we are assuming allows a world that is (toward the past) macroscopically like ours, in which I type "t" here, but does not allow one in which I type "q" in that same place. Remember, we are not concerned with the actual past history of the world in all its microscopic detail; that *does*, of course, determine the present including that typing of "t". We have set aside this traditional problem by adopting the perspective of freedom from the inside out. Instead we are here only concerned with whether there should be a physically possible world similar to actuality in some gross, macroscopic ways, and in which I (or my counterpart, if you like) types "q". How could it be the case that physics makes room for the one, but not for the other?

The harmony worry thus boils down to this: that our posited deterministic physics may allow vastly fewer possible worlds than we can imagine, so few that our normal conception of the could-have-done-otherwise is mistaken. And so few that what *seems* like good evidence for multiple choices in a given context (such as the evidence described above) is in fact not good evidence: there are very few physically possible worlds like ours, even though at least one of them (i.e., ours) happens to contain an abundance of misleading evidence in favor of freedom. Without having a genuinely adequate deterministic physics in hand to examine, there may be no way fully to resolve this doubt. There might be no way even if we did have the true physics in hand. But I am moved by the intuition that in *any* recognizable deterministic microphysics, there will be *so many* different micro-level world histories, there has to be more than enough scope for freedom as we normally conceive it.

3.3 Indeterministic microphysics

When we turn to considering freedom from the inside out under the assumption that an indeterministic microphysics holds in our world, things become simpler in one sense, and more complicated in others. Intuitively we expect the apparent challenge to freedom posed by such an underlying physics – always less clear-cut than the challenge from determinism – to dissolve more easily. Nevertheless, care is required in thinking through the possibilities under indeterminism.

Again we insist on downward causation, and the need for *past* history (at the micro-

level) to conform to the constraints set by the free choices of agents. Again we suppose that there is a past micro-state compatible with my typing "t" now, but also a macroscopically identical micro-state (which may or may not be different!) compatible with my typing "q" instead. But now the constraint is only that these micro-histories must be consistent with our merely probabilistic laws. Surely this is a looser set of constraints, and hence an easier context in which to maintain freedom?

Perhaps, but this does not follow immediately and trivially from the mere *idea* of an indeterministic microphysics. First of all, notice that in a formal sense determinism could fail (and indeterminism reign) without the challenge to free action changing significantly. Suppose, for example, that it remains the case that the state of the world (over the relevant region) a million years ago makes each and every one of our actions have a probability greater than 99.999%. I submit that this does not alter the force of the traditional incompatibilist argument that we are unfree very much. But things could be worse still; it might be that in fact our actions are all 100% necessitated by the past of 1 million years ago. Suppose that indeterminism holds only in this weak sense: once every 3 million years, an atom of hydrogen pops into existence at a random location in the universe; and this last happened 2.5 million years ago. Otherwise, events follow iron deterministic laws. This scenario posits what is, in some formal sense, an indeterministic world; but in essence things are just the same as under "pure" determinism.[17]

But recall that we are advocating freedom from the inside out. It is not necessary to maintain that we are only *loosely* constrained by the past, to maintain that we have freedom. Instead we simply maintain that the constraint goes the other way around. The past is (partly) constrained by our choices. How will this partial constraint play out under an indeterministic microphysics? Again, as above, it is not possible to make definitive pronouncements without having the physical laws before us (and it might be practically impossible even then). Plausibility considerations are the best we can do.

Prima facie it seems plausible that an indeterministic microphysics, which allows (by definition) multiple futures branching from a single past, should allow greater room for freedom than a deterministic microphysics. We intuitively picture a "branching tree" structure of possibilities, and think of the forks as corresponding to our free choices. The scenarios sketched above show us that we cannot automatically assume this is so. What matters, then, are the following questions: does our indeterministic microphysics allow various worlds corresponding to a variety of free actions we can undertake (in a given context) that all share an *identical* **or** *macroscopically identical* past? And does it do so for all of our free actions together, so that they harmonize appropriately?

[17]These brief remarks are meant to counteract a common tendency in the free will literature, that of conflating indeterminism with a sort of "anything goes" conclusion about what actions are physically possible given a fixed (micro- and macro-) past.

What is needed, then, is the same as in the case of determinism: a rich variety of physically possible worlds, so that we can take the actual history as one among many similar possible histories, whose actuality is explained (in part) by all our free choices. We need to be able to say that, generally, we *could* have done otherwise in the circumstances where we normally believe this; this might or might not imply that the past would have had to be different at the micro-level. The "might or might not" in the previous sentence is what distinguishes an indeterministic microphysics from a deterministic set of laws. Those who equate indeterminism with automatic room for freedom are assuming that these four words can be replaced with "would not". But this cannot be taken for granted.

What can be taken for granted is just the set of considerations developed above in section 3.2. An indeterministic microphysics might have a richer variety of worlds than a deterministic microphysics; but for all we know, it might not. To be the correct laws for *our* world, it must allow a great deal of variety – including the phenomena adduced in §3.2 as evidence that we should not worry about harmony problems. In the end, then, the situation seems to be the same as in the deterministic case.

4 Clarification of an Old Idea

Let me recap the main features of the notion of freedom from the inside out. We carefully distinguish the true story of the *physical* world as it is in itself, which is that of a block universe with only B-series time, from the world of everyday *experience* and *action*, which is wholly within A-series time. Physical determinism, if true at all, is true of the block universe with its B-series time, and implies no explanatory priority of the past over the future, or of future over past, or of the middle over the far past and future. It is therefore open to us to conceive of our actions as genuinely free, *properly* only explained by our desires, beliefs and intentions despite being *logically* determined by vast states of the world at other times.

While I have not seen this idea put forth in any modern discussion of free will and determinism, I must confess that I believe the first philosopher to advocate it was not me, but Kant. Kant famously defended a metaphysical picture that postulated a Newtonian/deterministic physical world, but also claimed that rational beings were *genuinely* the authors of their own free actions. How Kant thought he could reconcile these two theses is rarely discussed in a satisfactory way. A typical (and unsatisfactory) way of reading Kant's suggestions on this point is to read him as claiming that, in purely rational/intellectual terms, a person *qua* transcendental being should be considered the author of his/her own actions, at least for the purposes of praise and blame. But there are also cryptic comments about the whole of a person's life actions being but a single phenomenon, and a strong suggestion that the non-temporality of the noumenal world is what allows us to think of a person's will as the genuine source of their actions, despite determination by past events in the *phenomenal* (A-series) world. Here are some passages from the *Critique of Practical*

Reason:

> "… Repentance is entirely legitimate, because reason, when it is a question of the law of our intelligible existence (the moral law), acknowledges no temporal distinctions and only asks whether the event belongs to me as my act, and then morally connects it with the same feeling, whether the event occurs now or is long since past. For the sensuous life is but a single phenomenon in the view of an intelligible consciousness of its existence (the consciousness of freedom)… [and] must be judged not according to natural necessity which pertains to it as appearance but according to the absolute spontaneity of freedom" (LWB translation, p. 102).

> "… [despite the determination of a person's actions], we could nevertheless still assert that the man is free. For if we were capable of another view… i.e., if we were capable of an intellectual intuition of the same subject, we would then discover that the entire chain of appearances, with reference to that which concerns only the moral law, depends upon the spontaneity of the subject as a thing-in-itself, for the determination of which no physical explanation can be given." (LWB translation, p. 103)

In other words, Kant suggests, the agent as a noumenal being should be considered as the source and genuine explainer of his/her own free actions, even though *qua* physical things *in* time their actions are determined by earlier physical events.

Kant did not have McTaggart's distinction at his disposal. If we bring it to bear on Kant's metaphysical picture, we can clarify (and correct) that picture as follows. The block universe is the realm of things in themselves, i.e., the world *in itself*, not *as experienced by consciousness*. For Kant, "time" meant A-series time, and that is indeed restricted to conscious/rational experience. Physics, which does try to describe the world in itself (contra Kant's epistemic restrictions), needs only B-series time, i.e., a structure of relations among events that *underlies* and is partly isomorphic to A-series time. Rational agents can be understood as the ultimate explanatory sources of their own free actions[18]; the rest of the noumenal world , i.e., the rest of the block universe, must simply be such as to accommodate those actions. The only real mistake Kant made was in the locus of determinism: he thought it must be a feature of the world of experience, due to the necessary conditions of possible experience. In fact determinism is *no part* of our experience of the world, and if true at all, is only true at the subtle level of ultimate particles. Nevertheless Kant seemed to have the fundamental point right: when agents are conceived as "things in themselves" (i.e., as rational beings rather than as merely physical objects), their free actions are quite compatible with

[18]At this point one perhaps wants to hear more about the positive characterization of freedom that should accompany the negative picture (i.e., sketch of how the physical world leaves us *room* for freedom) developed above. I will not try to sketch or defend any positive account, but I am attracted to the basic idea we find in Kant: free action in the highest sense is action that springs not from mere desire, but rather from something intellectual, a concept of the good.

overall physical determinism, because those actions can be thought of as *outside* the time series (i.e., the A-series with its allegedly fixed past) and hence *not* unfree despite being "determined" by physical events lying to the past of them.

Whether or not this is really close to what Kant had in mind, I think it is what we should believe to be the case, if our world is causally complete. Free action and causal completeness *are* compatible after all, and not in the (arguably) weak sense offered by traditional forms of compatibilism. You have choices, and you make them. Because of determinism, your choices (like any events) place constraints on what the world's history can be. But the direction of determiniation (and, for most free actions, correct explanation) is *from* your choices *to* the ways the physical world can be – both toward the past and the future.

This picture of freedom from the inside out is more Idealistic than some will find comfortable. Take a God's-eye perspective on the block universe, and ask the question (Q): why are things as they are in it? A 21st-century materialist is comfortable with this sort of answer: "Well, you see, there was this Big Bang at the beginning, and after that things just sort of bump around in the ways permitted by the laws of nature, and that leads to the whole history." But this answer (a) is infected with the A-series view of time, (b) seems to presuppose an eliminativist picture of human thought and action, and (c) begs the question "Why was the Big Bang just so and not otherwise?" (a) is a mistake, (b) is at least dubious, and (c) is the lump under the carpet which, if you try to flatten it, leads to moves in all sorts of unpleasant directions (theology, Cosmic Anthropic Principles, and so on).

I prefer the picture that starts with what we *feel* so strongly that we really have: freedom to act in a variety of ways. This picture places some constraints – probably only very weak ones – on what an answer to this ultimate question (Q) can look like, if one is possible at all. It is Idealistic, in that the constraints involve giving rational agents priority over trivia such as the physical micro-state of vast regions of space-time (past or future). But this is a form of Idealism that most of us can learn to live with. Appropriately, it is McTaggart's distinction that helps us see that it is not nearly so strange as it may at first appear.

Bibliography

Dupré, John *The Disorder of Things*
Fischer, John *The Metaphysics of Free Will* (Blackwell, 1994)
Forrest, Peter "Backward Causation in Defence of Free Will", *MIND* 1985, pp. 210 - 217.
Horwich, Paul *Asymmetries in Time* (MIT Press, 1987).
Kant, Immanuel *Critique of Practical Reason*, edited and translated by Lewis White Beck (Maxwell Macmillan Publishing, 1993).
Price, Huw *Time's Arrow and Archimedes' Point* (Oxford University Press, 1996).
Russell, Bertrand "On the Notion of Cause" in *Mysticism and Logic* (Allen & Unwin 1908).
van Inwagen, Peter, "The Incompatibility of Free Will and Determinism", *Philosophical Studies* 27, pp. 185 - 99.

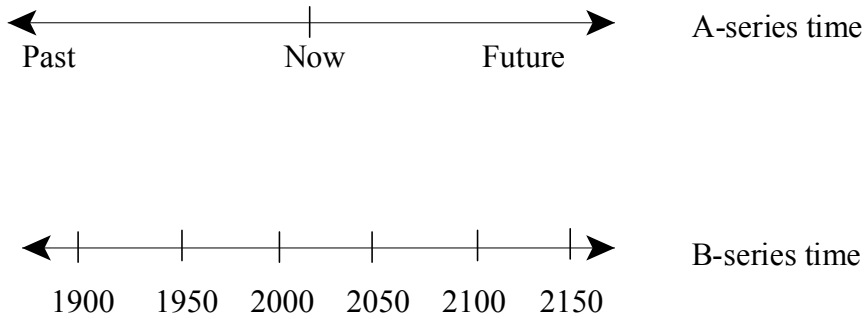


Figure 1: the two times

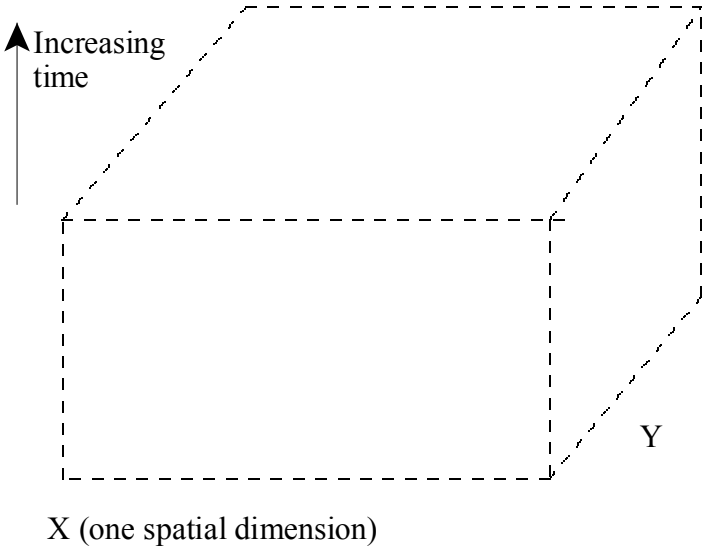


Figure 2: the block universe

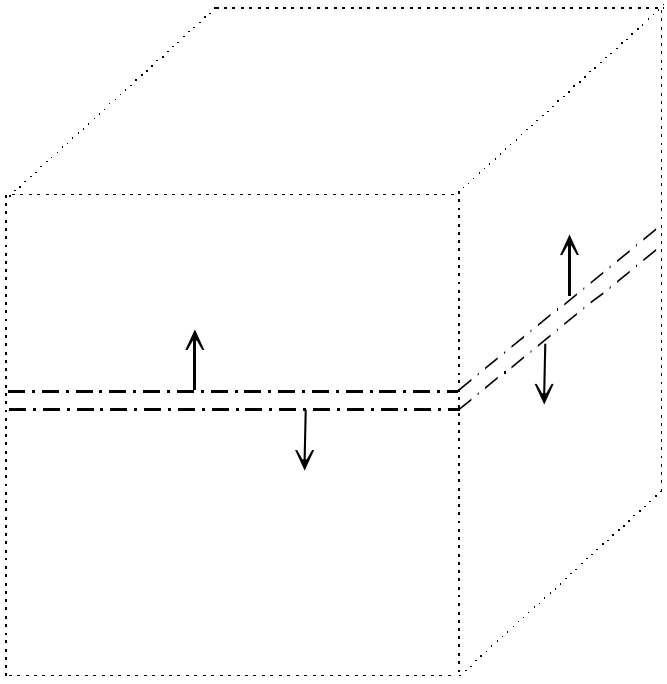


Figure 3: Freedom from the inside out. All human actions occur inside the sandwich of the block universe. Arrows indicate ***partial*** determination of physical events outside, by the actions within the sandwich.

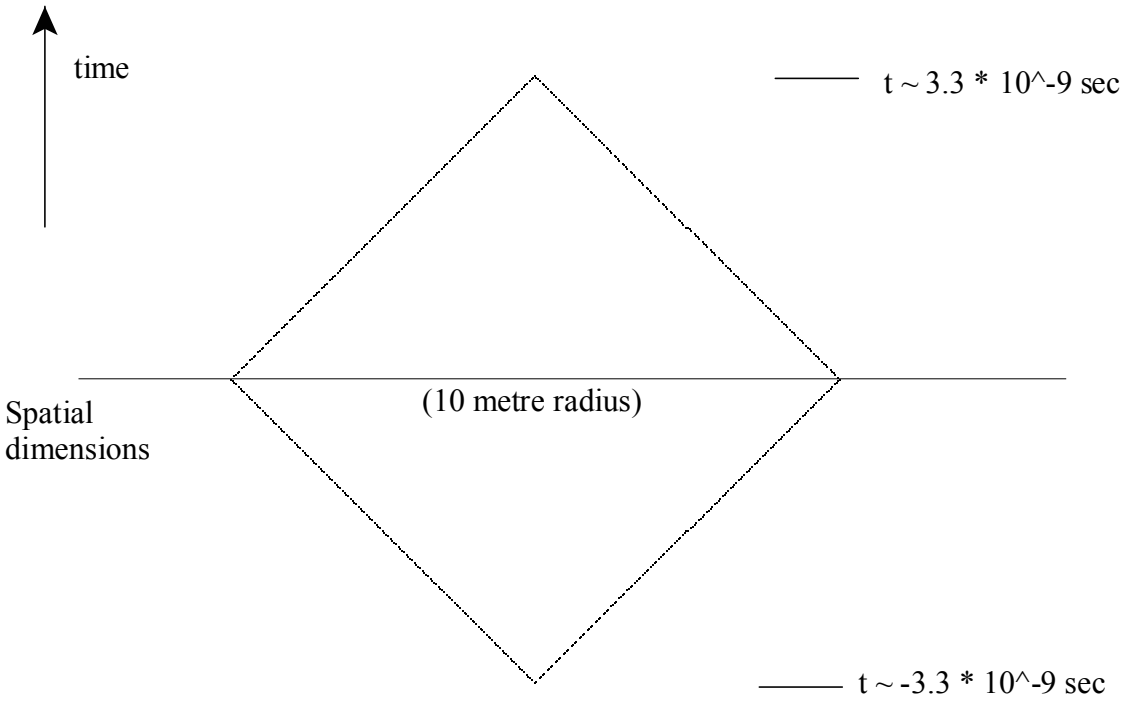


Fig. 4: space-time regions determined by events on hypersurface of 10m radius