Theoria, 2022, 37(1), 87-102
https://doi.org/10.1387/theoria.22841

THEORIA

ISSN 0495-4548 – eISSN 2171-679X

The fate of causal structure under time reversal

El destino de la estructura causal bajo inversiones temporales

Porter Williams*
Department of Philosophy, University of Southern California

ABSTRACT: What happens to the causal structure of a world when time is reversed? At first glance it seems there are two possible answers: the causal relations are reversed, or they are not. I argue that neither of these answers is correct: we should either deny that time-reversed worlds have causal relations at all, or deny that causal concepts developed in the actual world are reliable guides to the causal structure of time-reversed worlds. The first option is motivated by the instability under intervention of time-reversed dynamical evolutions. The second option is motivated by a recognition of how contingent structural features of the actual world shape, and license the application of, our causal concepts and reasoning strategies.

KEYWORDS: causality, causal structure; reversed time; causal concepts; intervention.

RESUMEN: *¿Qué sucede con la estructura causal de un mundo cuando se invierte el tiempo? A primera vista parece que hay dos posibles respuestas: o bien las relaciones causales se invierten, o bien no lo hacen. Defiendo que ninguna de estas respuestas es correcta: debemos o bien negar que los mundos con tiempo invertido tengan relaciones causales, o bien negar que los conceptos causales desarrollados en el mundo actual sean guías fiables para analizar la estructura causal de mundos con tiempo invertido. La primera opción está motivada por la inestabilidad bajo intervenciones de las evoluciones dinámicas con tiempo invertido. La segunda opción está motivada por el reconocimiento de cómo aspectos estructurales contingentes del mundo actual dan forma a, y legitiman la aplicación de, nuestros conceptos causales y estrategias inferenciales.*

PALABRAS CLAVE: causalidad; estructura causal; inversión temporal; conceptos causales; intervención.

*** *Correspondence to:*** Porter Williams. Department of Philosophy, University of Southern California, Los Angeles, CA 90089, USA. – porterwi@usc.edu – https://orcid.org/0000-0003-0242-4045

How to cite: Williams, Porter (2022). «The fate of causal structure under time reversal»; *Theoria. An International Journal for Theory, History and Foundations of Science,* 37(1), 87-102. (https://doi.org/10.1387/theoria.22841).

Received: 2021-05-26; Final version: 2022-01-31.

ISSN 0495-4548 - eISSN 2171-679X / © 2022 UPV/EHU

1. Introduction

Bertrand Russell famously declared that causal notions in fundamental physics were a "relic of a bygone age" (Russell, 1912). His most influential argument for this conclusion invoked the time-symmetric character of dynamical evolution in fundamental physics and has come to be called the *Directionality Argument*. The argument, in short, begins with the claim that the relationship between cause and effect is asymmetric in ways that cannot be grounded in time-symmetric dynamical laws. Since the dynamical laws of fundamental physics *are* time-symmetric, they cannot satisfactorily distinguish cause from effect: if $A \rightarrow B$ in one temporal direction, then $B \rightarrow A$ in the other temporal direction. Russell concludes that fundamental physical theories cannot ground causal relations.

Russell is far from alone in believing that reversing temporal order also reverses causal order. Of course, anyone who concludes with Russell that the Directionality Argument shows that causal relations cannot be grounded in fundamental physical theories certainly believes it. Even those who explicitly reject the conclusion of the Directionality Argument believe it on other grounds. For example, any advocate of transference theories, like (Salmon, 1984, 1994) or (Dowe, 2000), will certainly believe that reversing time reverses causal relations: a causal interaction between A and B *just is* the transfer of a conserved quantity, like energy-momentum, from A to B. The time-reverse of that dynamical evolution, of course, will involve the transfer of energy-momentum from B to A.

In a similar vein (Ney, 2009), explicitly endorses an account of "physical causation" that is just nomological determination—a temporally symmetric notion—and embraces the fact that

> ... if the universe does not have any fundamental, built-in temporal asymmetries, this seems to be what we are left with. There is still causation, because there is still physical determination. But the distinction between what is the cause and what is the effect may not be fundamental. (Ney, 2009, pp. 752-3)

Meanwhile (Price, 2007) endorses a quite different account of causation—a perspectival account according to which causal asymmetries are determined by our psychological perspective as deliberators—but motivates it by stating that if there were a region of the universe in which entropy was decreasing, "intelligent creatures would have a time-sense reversed relative to ours... For them, the causal arrow runs directly counter to the way it runs for us" (Price, 2007, p. 273). Indeed, Price thinks that his belief that causal order will be reversed in time-reversed worlds, combined with the time-symmetric character of the dynamical laws of fundamental physics, "provides something close to a trump card for perspectivalism" (Price, 2007, p. 269).[1]

And (Tooley, 1990, section 3.1.2) seems to think that *any* account of causation according to which causal relations supervene on non-causal facts and relations is committed to accepting that reversing temporal order will also reverse causal order.

[1] Price makes an important assumption about an anti-entropic universe: that such a universe does not require any more fine-tuning than an entropic one. As he says, "at least in the absence of any time-symmetry in the underlying physics, a fine-tuning required is the same in either temporal direction" (Price, 2007, p. 273). I think this assumption is false, as I will discuss in section 2.

I think this is mistaken: worlds with reversed temporal order should not be thought of as exhibiting reversed causal relations. To this extent I agree with (Farr, 2020). Farr points out that if causal relations in one temporal direction satisfy standard statistical desiderata, like the Causal Markov Condition, they will generally not satisfy those criteria in the reversed temporal order. On the basis of these asymmetries of statistical independence, he concludes that the causal structure of a world is invariant under time reversal:

> The issues of agency and the [Causal Markov Condition] lead to the same judgements about causal direction regardless of what one takes to be the underlying direction of time. This entails that any underlying time-reversal invariance of the microphysical description is beside the point; one may hold that there is a clear causal direction ... which is invariant under time reversal. (Farr, 2020, p. 197)

However, here I part ways with Farr as well: I do not think that reversing temporal order should be thought of as leaving causal ordering invariant. Instead, I think that the correct attitude is one of the following:

1. There are no causal relations *at all* in time-reversed worlds.
2. We have no epistemic warrant for judging that our concepts of causation are reliable guides to the nature, and presence, of causal relations in time-reversed worlds.

The two answers differ primarily in epistemic audacity. If we are audacious enough to apply to time-reversed worlds the concepts and strategies for causal reasoning that we have developed to help us navigate the actual world, then we should conclude that there are no causal relations at all in such worlds. Alternatively, upon recognizing how thoroughly foreign to us a time-reversed world is, we might decide that our actual-world standards for judging the nature and presence of causal relations are inapplicable. Ultimately, I think this second answer is the more plausible one: epistemic humility demands that we simply withhold judgment about the presence or absence of causal relations in time-reversed worlds, in light of the restricted scope of our frameworks for causal reasoning.

2. Reversing time

As (Farr & Reutlinger, 2013) have pointed out, the phrase "time-symmetric character" that appears in the Directionality Argument is ambiguous and can be precisified in two distinct ways:[2]

1. *Invertibility*: Given a state of the system $\mathcal{S}(t)$, the dynamics $\mathcal{D}_{\tau}$ of the physical theory determine both the state $\mathcal{S}(t+\tau)$ and the state $\mathcal{S}(t-\tau)$.
2. *Time-Reversibility*: Given a sequence of states of the system $\mathcal{S}(t_1)$, ..., $\mathcal{S}(t_n)$ related by the dynamics $\mathcal{D}_{t_n}$ and a time-reversal operator $\mathcal{R}$, the time-reversed sequence of states $\mathcal{RS}(-t_n)$, ..., $\mathcal{RS}(-t_1)$ is also related by the dynamics D_{-t_n}.

Both (Earman, 2002) and (Farr & Reutlinger, 2013) see *Invertibility* as the more fundamental sense in which the dynamics of a theory can exhibit a"time-symmetric character": Ear-

[2] A similar distinction was pointed out by (Earman, 2002), but was not applied to the Directionality Argument.

man because any classical or quantum system admitting a Hamiltonian formulation will satisfy *Invertibility* even if it doesn't satisfy *Time-Reversibility*, and Farr and Ruetlinger because one can understand the two dynamically allowed sequences of states secured by *Time-Reversibility* to be occurring in the same temporal direction.[3] I won't make much of the distinction: in my discussion below I will assume that both conditions are always satisfied. As is evident from the statement of *Time-Reversibility*, I will also adopt the standard interpretation of time reversal according to which time reversal includes a (perhaps non-trivial) operation on the state of a system in addition to changing the sign of the time coordinate.[4]

I will also assume that the dynamics of the theory are deterministic. This is sufficient to account for all "fundamental" dynamical evolutions of classical mechanics and for unitary evolution in quantum theories. For simplicity I will stick to the context of classical statistical mechanics, but see (Williams, 2022) for discussion of how some of the themes here apply to quantum states related by unitary dynamical evolution.

Consider a classical description of the behavior of an N-particle system whose dynamical evolution is governed by a Hamiltonian $\mathcal{H}$. This system could be anything: a gas in a box, a basset hound, the City of Pasadena, the entirety of the Pacific Ocean, or the observable universe as a whole. In all of those cases, the *microstate* $\mathcal{S}(t)$ of the system is determined by specifying the position and momentum of each particle:

$$\mathcal{S}(t) = (q_1(t), p_1(t), ..., q_N(t), p_N(t))$$

Dynamical evolutions of this system trace out continuous curves connecting different microstates $\mathcal{S}(t_1)$, ..., $\mathcal{S}(t_n)$ of the system. The space of all possible microstates of this system is called the *phase space* of the system.

We would also like to describe thermodynamic properties of the system. Roughly speaking, one accomplishes this by partitioning the phase space of the system into *sets* of microstates; we call these sets *macrostates*. All of the microstates contained in a particular macrostate describe a system that exhibits the same values for some set of specified thermodynamic properties like temperature, pressure, total magnetization, etc. It is the values of these thermodynamic properties that define the different macrostates. We can then assign an *entropy* to the system: the value of the entropy of the system at any instant is a function of the thermodynamic properties of the system at that instant. More precisely, the entropy assigned to a system at an instant is determined by the volume of phase space occupied by the macrostate in which the microstate $\mathcal{S}(t)$ of the system lies at that instant.[5]

---

[3] Farr and Ruetlinger draw on the same point made earlier by (Maudlin, 2007, chapter 4.2).

[4] A referee wanted justification for this. The appropriate understanding of time reversal has received a lot of philosophical attention in recent years (Albert, 2000; Callender, 2000; Earman, 2002; Malament, 2004; Arntzenius & Greaves, 2009; Roberts, 2017; Allori, 2019; Farr, 2020; Donoghue & Menezes, 2019, 2020; Callender, 2020; Struyve, 2020). The initial stimulation for much of this work were the arguments for a non-standard definition of time reversal by (Albert, 2000) and (Callender, 2000). For reasons compactly summarized in (Roberts, 2019), I remain partial to the traditional account and will adopt it throughout this paper.

[5] By *entropy* here I mean *Boltzmann entropy*. Nothing in the paper depends on this choice. For an illuminating and thorough discussion of entropies in classical and quantum mechanics, see (Goldstein *et al.,* 2020).

It has been known since the time of Boltzmann that the entropy of closed systems is overwhelmingly likely to increase over the course of their dynamical evolution: for essentially any Hamiltonian $\mathcal{H}$, essentially any non-equilibrium microstate $\mathcal{S}(t)$ will evolve to a later microstate $\mathcal{S}(t + \tau)$ of higher entropy. However, since Boltzmann's time it has also been known that for essentially any Hamiltonian $\mathcal{H}$, essentially any non-equilibrium microstate $\mathcal{S}(t)$ will also evolve to an *earlier* microstate $\mathcal{S}(t - \tau)$ of higher entropy: as (Albert, 2000, p. 77) puts it, "the overwhelming majority of the trajectories passing through any particular non-maximal-entropy [macrostate] must just then be in the process of *turning around*." This is troubling: the claim that essentially any non-equilibrium microstate $\mathcal{S}(t)$ represents a local entropy *minimum* along essentially every dynamical trajectory that passes through it conflicts directly with our memories of the past: isolated gases were not warmer in the past than they are in the present, my eggs were not more scrambled, my hair (sadly) not less grey, and so on.

However, "essentially any microstate" does not mean "every microstate": included in the set of microstates that make up any macrostate, there will be a subset of microstates $\mathcal{S}_{ab}(t)$ that, under the influence of $\mathcal{H}$, evolve to higher entropy states in one temporal direction and to *lower* entropy states in the other temporal direction. This subset of "abnormal" microstates states $\mathcal{S}_{ab}(t)$ inhabits a fantastically miniscule portion of the phase space volume occupied by the macrostate in question. Furthermore, this miniscule volume is itself scattered in tiny, geometrically non-uniform clusters and filaments throughout the volume of phase space associated with the macrostate in question; see Figure 1. Given their sparseness, it would seem fantastically unlikely that the present state of the actual world is such a state; nevertheless, all evidence indicates that it is. Providing an account of the foundations of statistical mechanics according to which such a history of the world is not fantastically unlikely, but in fact *highly probable*, has generated a host of conceptual difficulties in the foundations of statistical mechanics (Sklar, 1993; Albert, 2000; Uffink, 2007).

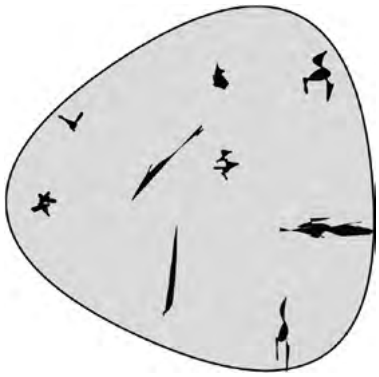

Figure 1

The set of abnormal microstates $\mathcal{S}_{ab}$ distributed throughout a macrostate.
(Not drawn even close to scale.)

However one secures such an account, the result is that the universe as a whole—and essentially any closed system within that universe—dynamically evolves from microstates

of lower entropy to microstates of higher entropy. Call worlds like this *entropic*. The actual world is entropic. The time-reverse of an entropic world is a world in which microstates of higher entropy dynamically evolve into microstates of lower entropy. Call worlds like this *anti-entropic*. The time-reverse of the dynamical history of any entropic world corresponds to the dynamical history of an anti-entropic world.

A characteristic feature of entropic worlds is that their identity *as* entropic worlds is very stable. Suppose that any physical system, up to and including the universe as a whole, is in a particular non-equilibrium microstate $\mathcal{S}(t)$ that will evolve, under the influence of the Hamiltonian $\mathcal{H}$, into a microstate $\mathcal{S}(t + \tau)$ of higher entropy. As stated above, essentially any *alternative* microstate $\mathcal{S}'(t)$ that lies in the same non-equilibrium macrostate as $\mathcal{S}(t)$ will also evolve, under the influence of the same Hamiltonian $\mathcal{H}$, into a microstate $\mathcal{S}'(t + \tau)$ of higher entropy. No particular coordination, or fine-tuning, of the microstate to the particular Hamiltonian $\mathcal{H}$ is required to secure entropic dynamical evolution.

It is worth lingering on this point for a moment because elucidating the type of fine-tuning that is, and is not, required to secure entropic dynamical evolution brings out an important difference between entropic and anti-entropic worlds.[6] Many proposals for ensuring that our world lies on a dynamical trajectory along which low entropy microstates evolve into high entropy microstates posulate something like the Past Hypothesis: that the initial microstate of the universe had very low entropy (Feynman, 1965, chapter 5; Albert, 2000; Wallace, 2011). Macrostates with low entropy occupy a small volume of the phase space of any physical system and the universe itself is no exception, so these proposals require a certain type of fine-tuning: the initial state of the universe had to be "special" in the sense that it had to be selected from a small volume of phase space. But that's it: to ensure entropic dynamical evolution, there is no need to choose a *particular* microstate that is fine-tuned to the *particular* Hamiltonian governing the system. In fact, the specific form of the dynamics isn't all that important: the set of microstates that occupy the low-entropy macrostate for the Hamiltonian $\mathcal{H}$ will be essentially the same for a reasonably large neighborhood of different Hamiltonians $\mathcal{H}'$.

In anti-entropic worlds, the situation is importantly different. A characteristic feature of anti-entropic worlds is that their identity *as* anti-entropic worlds is extremely *fragile*. The anti-entropic dynamical evolution of a physical system requires a fantastically precise dynamical coordination between the constituent particles of the system: all the particles in my kitchen coordinate to unscramble eggs and launch them back into their shells, the particles in my coffee and mug and local environment cooperate to unmix my coffee and cream, stomach acids conspire to reassemble digested foodstuffs, and so on. If even a few particles fail to play their role, the anti-entropic dynamical evolution falls apart. A few displaced molecules in my frying pan collide with, and disrupt, the carefully coordinated trajectories of other molecules, those molecules fail to displace the appropriate molecules in my scrambled eggs and/or the surrounding air molecules, and so on.[7]

[6] Much of my discussion about the differences in fine-tuning is indebted to (Maudlin, 2007, chapter 4.4).

[7] Similar observations about the fragility of anti-entropic evolution are invoked by (Elga, 2001) to present a problem for David Lewis's account of the asymmetry of counterfactual dependence. See, in particular, Elga's description of the anti-entropic dynamical evolution of a fried egg.

We can be slightly more precise about this. If a microstate $\mathcal{S}(t)$ lies on a dynamical trajectory along which high entropy microstates evolve into low entropy microstates, then essentially any microstate $\mathcal{S}^{\epsilon}(t)$ obtained as a small perturbation of $\mathcal{S}(t)$ will *not* lie along such a dynamical trajectory. That is, if a microstate $\mathcal{S}(t)$ is in the set $\mathcal{S}_{ab}(t)$ of "abnormal" microstates that undergo anti-entropic dynamical evolution under the influence of a particular Hamiltonian $\mathcal{H}$, then the fact that $\mathcal{S}_{ab}(t)$ occupies such a fantastically tiny, and scattered, volume within any macrostate means that small perturbations of $\mathcal{S}(t)$ will almost certainly *not* be in the set $\mathcal{S}_{ab}(t)$. For example, if the microstate

$$\mathcal{S}(t) = (q_1(t), p_1(t), ..., q_N(t), p_N(t))$$

lies in the macrostate Γ *and* the subset $\mathcal{S}_{ab}(t)$ of "abnormal" microstates, then the microstate $\mathcal{S}^{\epsilon}(t)$ that results from a small perturbation of the positions and/or momenta of some subset of the particles

$$\mathcal{S}^{\epsilon}(t) = (q_1(t), p_1(t), ..., q_i(t) + \epsilon, p_i(t) + \epsilon, ..., q_m(t) + \epsilon, p_m(t) + \epsilon, ..., q_N(t), p_N(t))$$

will still lie in the macrostate Γ but will *not* lie in $\mathcal{S}_{ab}(t)$. This means that, evolving under the influence of the Hamiltonian $\mathcal{H}$, the perturbed microstate $\mathcal{S}^{\epsilon}(t)$ will generate an *entropic* dynamical evolution. This is one sense in which anti-entropic worlds are fragile: at any instant t, small perturbations of the microstate of that world will transform it into an entropic world, i.e. shift it onto a dynamical trajectory along which low entropy microstates evolve into higher entropy microstates.[8] I will return to this in section 3.

There is a second type of fine-tuning required in anti-entropic worlds: the microstate generating the anti-entropic evolution has to be delicately tuned to the *particular* Hamiltonian $\mathcal{H}$ that determines that dynamical evolution of the system. Physically, this is because the coordination between all of the particles of a system required for an anti-entropic dynamical evolution depends extremely sensitively on the particular Hamiltonian governing the system. If a microstate $\mathcal{S}(t)$ would generate an anti-entropic evolution under the influence of $\mathcal{H}$, then that same microstate will generate an *entropic* evolution under the influence of the Hamiltonian $\mathcal{H}^{\epsilon}$ obtained from $\mathcal{H}$ by some tiny modifications—for example, to the range or strength of interactions. Those modifications will (among other things) result in slightly different scattering angles from particle interactions which, in turn, will disrupt the careful coordination of particles secured by $\mathcal{H}$ and progressively wipe out the delicate correlations that are needed between all of the particles at every instant t to generate anti-entropic dynamical evolution. Formally, it is because the meaning of "abnormal" in the definition of the set of abnormal microstates $\mathcal{S}_{ab}(t)$ is *dynamical* abnormality: it is the set of microstates that undergo abnormal dynamical evolution, i.e. evolve from higher entropy microstates to lower entropy microstates. Modifying the Hamiltonian thus completely redefines the set of "abnormal" states $\mathcal{S}_{ab}(t)$.[9]

---

[8] This fact is exploited in (Albert, 2000, chapter 9) to argue that if a spontaneous collapse theory like GRW is the correct description of quantum behavior, then anti-entropic dynamical evolution is essentially impossible: the spontaneous localizations will take any state initially in the set $\mathcal{S}_{ab}(t)$ and very rapidly perturb it into a state that lies outside that set.

[9] The same point is made by (Maudlin, 2007, p. 132-33).

The fine-tuning in an anti-entropic world is different not only in degree, but also in kind, from the fine-tuning in entropic worlds associated with proposals like the Past Hypothesis. In entropic worlds one can randomly pick a microstate from a low-entropy macrostate without knowing much at all about the particular Hamiltonian $\mathcal{H}$ determining the system's dynamical evolution, then show that those dynamics will evolve that low-entropy microstate into a high-entropy microstate. The initial microstate is "special" in the sense that it comes from a macrostate that occupies a small volume in the system's phase space, but securing entropic dynamical evolution does not *additionally* require that "special" microstate to contain strong correlations between the positions and/or momenta of particles with no antecedent causal connection, nor does picking a "special" microstate that will generate entropic dynamical evolution depend sensitively on the details of a particular $\mathcal{H}$. In the terminology of (Woodward, 2022), both Causal to Statistical Independence (CSI) and Variable/Relationship Independence (VRI) are satisfied in entropic worlds. In anti-entropic worlds, this is no longer true: the initial microstate describes strong correlations between the positions and/or momenta of particles with no antecedent causal connection, and those correlations must be very precisely tailored to the particular Hamiltonian $\mathcal{H}$ determining the dynamical evolution of the system. In other words, both CSI and VRI fail in anti-entropic worlds.

3. Causation reversal?

What does any of this have to do with time reversal and causal ordering? The answer, in short, is that according to any difference-making account of causation—any account on which causal claims entail a claim about how an effect would be altered by local alterations of the cause—there are no causal relations in anti-entropic worlds. This includes interventionst accounts of causation, accounts of causation according to which causal claims reduce to (or non-reductively entail) claims about counterfactual dependence, and so on. If causation is a difference-making relation then inverting the temporal ordering of the states of an entropic world does not result in a world with inverted causal relations: it results in a world with no causal relations at all.

More precisely, the following two claims are inconsistent:

1. At least some causal claims, like *temperature causes pressure* or *smoking causes lung cancer*, are meaningful in world ω.
2. Closed systems in ω (including ω itself) dynamically evolve from states of higher entropy to states of lower entropy.

The argument for the inconsistency is simple, but first recall a couple elementary features of interventionist accounts of causation.[10] Claims like *temperature causes pressure* mean "there is some intervention that can be performed on the *temperature* of a physical system that is systematically correlated with a change in the *pressure* of that system." And to *intervene* on a variable $\mathcal{V}$ is to bring that variable, and that variable alone, under the total control

[10] I will adopt interventionist language from here on out, but the argument would go through just as well with any other difference-making account.

(or partial control, if one considers soft interventions) of the investigator while holding all other variables fixed. If there is no act one can perform—even in principle—to change the value of $\mathcal{V}$ without *that act*, and not the resulting change in $\mathcal{V}$, also changing the value of a distinct variable $\mathcal{Z}$, then one cannot intervene on $\mathcal{V}$. For example, if there was no act that one could perform to change the net magnetization of an iron bar that didn't also change its temperature, or the air temperature in the lab as a whole, or the net magnetization of the iron bar in the lab down the hall, or ... then, in such a world, one could not intervene on the net magnetization of an iron bar.[11]

Here, then, is the argument that (1) and (2) are inconsistent. We know from the discussion in section 2 that anti-entropic dynamical trajectories are extremely fragile: if a microstate $\mathcal{S}(t)$ generates an anti-entropic dynamical trajectory under the influence of a Hamiltonian $\mathcal{H}$, then states $\mathcal{S}^{\epsilon}(t)$ obtained by tiny perturbations of $\mathcal{S}(t)$ almost surely will not. I want to emphasize that by "tiny perturbations" I mean perturbations that are completely physically unnatural, let alone attainable: changing by a negligible amount the positions and/or momenta of a few random particles in large system, like a dilute gas, a frying pan, a few cells in a human body, downtown Los Angeles, and so on. Such interventions would require controlling the state of every particle in the system to a physically unattainable degree of precision, but that's not really the issue. Even if they were physically attainable, there is no way to know *which* few particles in the system, if any, one could subject to such a tiny perturbation at a given instant without disrupting the anti-entropic dynamical evolution: that is, without perturbing the system in the state $\mathcal{S}(t)$ into a state $\mathcal{S}^{\epsilon}(t)$ that lies outside any of the tiny, geometrically non-uniform regions of phase space that make up $\mathcal{S}_{ab}(t)$. And even if one could know, and could perform such a precise intervention, these would not be interventions on anything remotely like meaningful physical properties of a system: there is no physically interesting property corresponding to the small and scattered collection of particles in downtown Los Angeles that one could minimally perturb without disrupting the anti-entropic dynamical evolution of the city as a whole.

Genuine collective physical properties, like a nucleotide sequence in an mRNA molecule, the temperature of an iron bar, or the amount of cigarette tar in one's lungs, are associated with the positions and/or momenta of *enormous* numbers of individual particles. And interventions on those propeties, like mRNA editing, lowering the temperature of an iron bar, or smoking one pack a day rather than two, involve large changes in the positions and momenta of enormous numbers of particles. Those kinds of interventions will inescapably take a state $\mathcal{S}(t)$ that lies in $\mathcal{S}_{ab}(t)$ and produce a state that lies outside of $\mathcal{S}_{ab}(t)$—a state that will produce an *entropic* dynamical evolution under the influence of $\mathcal{H}$. Local interventions on physical systems cannot be performed without transforming an anti-entropic world into an entropic one.[12]

---

[11] Of course, in the real world one *can* intervene on the net magnetization of an iron bar precisely by changing its temperature: cool it below its critical temperature. This brings out the difference nicely: one is acting to intervene on the temperature of the bar and it is the change in the temperature that *brings about* a change in its net magnetization. It is *not* whichever specific action one took to lower the temperature that *itself* brought about the change in net magnetization in the iron bar.

[12] The qualifier "local" is important. One could perform interventions on $\mathcal{S}(t)$ while maintaining anti-entropic dynamical evolution, but those "interventions" would have to be spectacularly non-local. Such an intervention would have to *instantaneously* alter the positions and/or momenta of particles

It may seem incredible that that such small interventions can end up reversing the entropy gradient of the universe, i.e. turning an anti-entropic world into an entropic world. It is clarifying to think about the physical process by which this transformation happens. Suppose that the microstate of the universe is $\mathcal{S}(t)$ and that this microstate, under the influence of a Hamiltonian $\mathcal{H}$, will generate anti-entropic dynamical evolution. Among the countless physical states of affairs described by the microstate $\mathcal{S}(t)$, my office windows are currently shut. An intervention on the state of the window—opening it—will spatially translate the positions of all the particles in the window by some non-trivial amount. The delicate choreography of particle collisions required to secure anti-entropic dynamical evolution will be disrupted: many collisions between air molecules, electromagnetic radiation, dust, molecules in the glass, etc. that needed to take place will not, while many others that would not have occurred now will. The delicate correlations between the positions and momenta of particles in the vicinity of my office window will be washed out by these collisions, resulting in an *entropic* dynamical evolution. On short time scales, only the collection of particles in a fairly small spatiotemporal region around my office window will undergo entropic evolution; the rest of the universe will continue to evolve anti-entropically.[13] However, at the boundary the particles in this spatiotemporal region will interact with the particles outside of it, disrupting their carefully choreographed dynamical evolutions and thereby expanding the size of the spacetime region in which particles undergo entropic evolution. After enough time, the originally anti-entropic world ω will have become a world in which every closed system (including the universe itself) undergoes *entropic* dynamical evolution. All because at some earlier time I opened my office window.

That is the entirety of the argument that (1) and (2) are inconsistent. One cannot perform interventions in anti-entropic worlds without turning them into entropic worlds. If causal claims like *temperature causes pressure* are understood as claims about what would happen to *pressure* under interventions on *temperature* then—insofar as interventions are incompatible with anti-entropic dynamical evolution—causal claims cannot be meaningful in anti-entropic worlds. Time-reversing an entropic universe does not produce a universe with reversed causal relations; it produces a universe with no causal relations at all.[14]

---

not only in the spatiotemporal vicinity of the intervention, but also in other causally disconnected spacetime regions. The result of a such an anti-entropic-evolution-preserving non-local intervention would be to remove $\mathcal{S}(t)$ from one of the "abnormal" regions in Figure 1 but to modify it in such a way as to produce a state that lies in one of the *other* "abnormal" regions. It is a non-starter to argue that the possibility of "interventions" like this would somehow save the day for meaningful causal relations in anti-entropic worlds. For starters, they wouldn't count as interventions in the technical sense at all. (For discussion of a few of the difficulties facing non-local "interventions" like this, see (Hausman & Woodward, 1999).)

[13] In (Elga, 2001) these spatiotemporal regions of entropic dynamical evolution are called "infected regions", although he puts the existence of such regions to somewhat different philosophical ends.

[14] A referee asks whether it makes sense to speak of interventions at all in an anti-entropic world, i.e. whether interventions themselves are necessarily entropy increasing. There are two things to say about this. The first is that an intervention, in the technical sense, is just a change in the value of a variable that has to satisfy certain other conditions (see (Woodward, 2003, chapter 3.1) or (Pearl, 2009, chap-

4. Conclusion

I want to consider two possible avenues of response. The first attempts to save the idea that there are causal relations in anti-entropic worlds. The second recognizes the essential role that contingent structure of the actual world plays in the development and application of our causal concepts and reasoning strategies and, in doing so, touches on some of the central ideas of (Woodward, 2022).

I have employed a difference-making notion of causation throughout. One could abandon such a notion in favor of one according to which, for example, causation is a matter of nomological determination (Ney, 2009) or the transfer of conserved quantities (Salmon, 1984, 1994; Dowe, 2000). These approaches to causation face severe difficulties (Hausman, 1998, chapter 1; Hausman, 2002; Glynn, 2013; Paul & Hall, 2013) and the fact that they would allow one to understand time reversal as entailing causation reversal does not seem to me to nearly outweigh those difficulties.[15] In fact, I think one philosophical consequence of the methods for inferring causal direction discussed in (Woodward, 2022) is to pose yet another difficulty for such approaches. By employing statistical independence conditions like CSI and VRI, one can ground a causal asymmetry between states of classical or quantum systems related by time-symmetric dynamical laws (Janzing *et al.,* 2016; Williams, 2022). As illustrated above, there is nothing necessarily emergent or non-fundamental about these statistical asymmetries. In classical statistical mechanics, they arise between *microstates* whose evolution is described by a Hamiltonian $\mathcal{H}$; if this does not satisfy the requirement that they occur in a theory "intended to describe the mechanisms of microphysical interactions" (Ney, 2009, p. 749), then nothing does. By reducing causation to a relation of nomological determination, such accounts indulge in what (Woodward, 2022, section 11) calls the "cause-in-laws" picture. As a result, such accounts are insensitive to the causal information encoded in the statistical asymmetries that exist between states related by fundamental dynamical laws. Such accounts face the burden of explaining why one can reliably infer causal direction on the basis of such statistical asymmetries if causation really is just nomological determination, i.e. if all of the causally relevant information is contained in the laws alone.

Alternatively, one could retain a difference-making account of causation and argue that the causal ordering of an entropic world remains fixed under time reversal. One might go about this strategy in a couple different ways. If one adopts a standard understanding of the time reversal operation, then a sequence of states $\mathcal{S}(t_1)$, ..., $\mathcal{S}(t_n)$ related by the dynamics $\mathcal{D}_{t_n}$ and the sequence of time-reversed states $\mathcal{RS}(-t_n)$, ..., $\mathcal{RS}(-t_1)$ also related by $\mathcal{D}_{-t_n}$ de-

ter 3), for example); it needn't be actually carried out by a human agent. Understood in that way, there are certainly interventions that do not themselves increase entropy: any change to the microstate of a system that does not change its macrostate. How large this class of interventions is will depend on specific details of the physical system and the particular macrostate in question. (Note that this is different than whether the intervention changes the future dynamical evolution of the microstate. As discussed above, it almost certainly will.) The second thing to say is that if interventions were impossible in anti-entropic worlds, then so much the worse for the idea that those worlds contain causal relations.

[15] It is notable that such accounts of causation fail to satisfy any of the quite different sets of criteria for evaluating accounts of causation proposed by (Hausman, 1998, chapter 1; Paul & Hall, 2013, chapter 2, and Woodward, 2014, respectively).

scribe, in general, distinct physical processes. Suppose that one takes the direction of causation to be the causal ordering that satisfies independence conditions like CSI and VRI; this is the same in an entropic world and its time-reversed, anti-entropic partner, so the direction of causation is fixed under time reversal. This entails that the direction of time and the direction of causation systematically come apart in anti-entropic worlds: effects will almost always precede their causes. Such an account retains meaningful causal relations in anti-entropic worlds, but at the expense of severing any connection between causal and temporal ordering. This is a huge conceptual cost: that connection is central to the everyday and scientific notion(s) of causation whose explication is, at least in part, the aim of providing a philosophical account of causation in the first place.

An alternative version of this strategy is pursued in (Farr, 2020). Farr's particular proposal aims to avoid this conceptual cost by adopting a non-standard treatment of time reversal. The reason Farr adopts this non-standard treatment is that it allows him to identify a world and its time-reverse as alternate descriptions of one and the same world; on this non-standard account, the state $\mathcal{S}(t)$ and its time-reverse $\mathcal{RS}(-t)$ are simply alternate descriptions of a single physical state of affairs. That means that $\mathcal{S}(t_1)$, ..., $\mathcal{S}(t_n)$ and $\mathcal{RS}(-t_n)$, ..., $\mathcal{RS}(-t_1)$ are simply two alternative descriptions of a *single* sequence of instanteous physical states of affairs. The correct causal ordering of that single sequence of physical states of affairs is determined by statistical independence conditions like CSI and VRI. This causal ordering picks out one of the two alternative descriptions $\mathcal{S}(t_1)$, ... or $\mathcal{RS}(-t_n)$, ... as the correct one, and thereby determines the direction of time.

The problem with this alternative version of the strategy hinges on its treatment of time reversal. Many physical quantities change under time reversal: the spin of an electron, the direction of a magnetic field, the momentum of a physical system, and so on. (As described by many of the references in footnote 4, there is good physical reason to require this.) To avoid the conclusion that a sequence of states and its time-reverse describe distinct physical processes, Farr needs to deny physical significance to properties that are not invariant under time reversal; otherwise an instantaneous state and its time-reverse could not simply be alternate descriptions of a single physical state of affairs. This requires Farr to claim that properties that change under time reversal, like momentum, the direction of a magnetic field, the spin of a quantum system, etc. "are either (i) not causal, or (ii) not genuine properties of instantaneous states" (Farr, 2020, p. 201).[16] A full examination of the challenges faced by such a proposal is largely orthogonal to my main focus here, but for present purposes I will say the following. Abandoning the standard interpretation of properties like spin and magnetic fields as genuine, causally efficacious physical properties is (i) *ad hoc*—the only motivation is that it is required to interpet $\mathcal{S}(t)$ and $\mathcal{RS}(-t)$ as describing a single physical state of affairs, which is itself only motivated by Farr's goal of maintaining a fixed causal direction under time reversal—and (ii) seems to have enormous (but unaddressed in (Farr, 2020)) consequences for the explanatory resources of our physical

[16] See also: "[time reversed]-twins differ only in terms of notation: they represent a single possible world, and hence notation that varies under time reversal (such as the direction of velocities) should not be taken to represent a property of the target system" (Farr, 2020, p. 191) and "any quantities that differ between [time reversed]-twins (such as instantaneous velocity, spin, and so on, as discussed above) can be considered descriptive artefacts that equally correspond to a single time-direction-independent... state of affairs" (Farr, 2020, fn. 18).

theories, apparently invalidating the myriad explanations that invoke these "causally inefficacious" quantities.[17]

I think the prospects for retaining meaningful causal claims in anti-entropic worlds are dim. That said, one might start to wonder what is really at stake here. Such worlds are fantastically foreign to us: eggs unscramble; photons are emitted from our retinas with wavelengths perfectly matched to their incident surfaces; food is reconstituted in stomachs, regurgitated, and spontaneously recombines; previously dead organic matter is reanimated and undergoes anti-aging; and so on. Life in such worlds defies imagination.[18] So there are no causal relations in such worlds: who cares?

This brings me to the second avenue of response. Down this second avenue lies an epistemically modest attitude: anti-entropic worlds are so different from the actual world that we have no epistemic warrant for judging the presence *or* absence of causal relations in those worlds based on attempts to apply causal concepts developed to reason and navigate in the actual world.

In particular, the wordly infrastructure that supports causal reasoning in the actual world—for example, the widespread satisfaction of independence conditions like CSI and VRI—is entirely absent in anti-entropic worlds.[19] This presents us with a dramatic mismatch between the worldly infrastructure of anti-entropic worlds and contexts in which our real-world causal concepts and reasoning strategies can be reasonably applied. This is because the concepts and strategies that we have developed for causal reasoning in the actual world are built to exploit precisely the worldly infrastructure that is absent in anti-entropic worlds. Our causal concepts and strategies are *designed* to be applied to worlds that share certain structural features with ours; in particular, worlds in which CSI and VRI are generically satisfied. Here we make contact with Woodward (2022) whose point, in part, is to remind us that our causal concepts have been developed by human beings—over a long evolutionary history—for the purpose of exploiting certain pervasive structural features of the world so they can achieve certain pragmatic and intellectual goals.[20] These structural features of the world include, among other things, various independence conditions like CSI and VRI that allow for local interventions and modular reasoning.[21] *When those structural features are present*, the causal concepts that we have developed to exploit those structural features enable us to make reliable causal judgments.

However, when confronted with a world in which those structural features are widely or uniformly absent—for example, worlds in which CSI and VRI fail dramatically—then epistemic honesty demands we recognize that we are at a loss. We should not expect the causal concepts and strategies that we have developed to navigate the actual world to be applicable: the worldly infrastructure that licenses their use and supports their successful ap-

---

[17] Whether quantities that include time derivatives in their definition, like momentum, are properties of instantaneous states is a more subtle question; see (Albert, 2000; Arntzenius, 2000; Smith, 2003).

[18] Indeed, I find the argument in (Maudlin, 2007, chapter 4.3) that beings in such worlds would not have conscious experiences at all fairly compelling.

[19] This raises an interesting question about the precise connection between the satisfaction of CSI and VRI and the fact that the actual world is entropic.

[20] See also (Woodward, 2014) for how this fact should impact the way we approach theorizing about causation.

[21] See also (Hausman & Woodward, 1999; Cartwright, 2002; Hausman & Woodward, 2004).

plication is not present. If such a world has causal structure at all, our methods are incapable of diagnosing it. Intelligent beings in such a world, presuming they could exist and that their survival (like ours) would depend on successfully exploiting various dependence relationships in that world, would inevitably develop a radically different set of concepts and strategies for doing so. Only hubris could suggest that we can apply our own parochial set of causal concepts to form reliable judgments about the presence or absence of causal relations in such a world.

Some will be inclined to think that by tying an account of causation to contingent features of the actual world so tightly as to restrict its applicability in this way, one is exhibiting something like a lack of philosophical ambition.[22] I disagree. Emphasizing the important role that independence conditions like CSI and VRI have played in shaping our causal concepts and supporting successful causal reasoning is, I think, an essential aspect of any naturalistic analysis of causation. Indeed, any such naturalism ought to be built on a foundation that, among other things, includes an acceptance of the contingency of our causal concepts and strategies for formulating successful causal judgments, and of the pragmatic forces driving their developmental history. An approach to analyzing causation that focuses on a detailed accounting of the worldly infrastructure that shapes and supports our causal reasoning strategies promises to yield a richly informative account of causation, but one that can be sensibly applied to a comparatively small volume of the space of possible worlds. Focusing on how contingent structural features of the actual world ground our causal concepts is not a widely adopted understanding of what it means to be engaged in providing an account of the metaphysics of causation, but I think it offers a promising path for those interested in naturalistic analyses of our causal concepts and reasoning strategies.

Acknowledgments

For helpful discussions and comments I would like to thank David Albert, Bob Batterman, Naftali Weinberger, and especially Jim Woodward.

REFERENCES

Albert, D. Z. (2000). *Time and Chance*. Harvard: Harvard University Press.

Allori, V. (2019). Quantum mechanics, time and ontology. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 66, 145-154.

Arntzenius, F. (2000). Are there really instantaneous velocities? *The Monist*, 83(2), 187-208.

Arntzenius, F. & Greaves, H. (2009). Time reversal in classical electromagnetism. *The British Journal for the Philosophy of Science*, 60(3), 557-584.

Callender, C. (2000). Is time 'handed' in a quantum world? *Proceedings of the Aristotelian Society*, 100(1), 247-269.

Callender, C. (2020). Quantum mechanics: keeping it real? *philsci-archive preprint: 17701*.

Cartwright, N. (2002). Against modularity, the causal Markov condition, and any link between the two: Comments on Hausman and Woodward. *The British Journal for the Philosophy of Science*, 53(3), 411-453.

[22] For example, in (Paul & Hall, 2013, section 3.2) this is sufficient to make one an "ontological wimp".

Donoghue, J. & Menezes, G. (2019). Arrow of causality and quantum gravity. *Physical Review Letters*, 123(17), 171601.

Donoghue, J. & Menezes, G. (2020). Quantum causality determines the arrow of time. *arXiv preprint arXiv:2003.09047*.

Dowe, P. (2000). *Physical causation*. Cambridge: Cambridge University Press.

Earman, J. (2002). What time reversal invariance is and why it matters. *International Studies in the Philosphy of Science*, 16(3), 245-264.

Elga, A. (2001). Statistical mechanics and the asymmetry of counterfactual dependence. *Philosophy of Science*, 68(S3), S313-S324.

Farr, M. (2020). Causation and time reversal. *The British Journal for the Philosophy of Science*, 71(1), 177-204.

Farr, M. & Reutlinger, A. (2013). A relic of a bygone age? causation, time symmetry and the directionality argument. *Erkenntnis*, 78(2), 215-235.

Feynman, R. (1965). *The Character of Physical Law*. Cambridge: MIT Press.

Glynn, L. (2013). Causal foundationalism, physical causation, and difference-making. *Synthese*, 190(6), 1017-1037.

Goldstein, S., Lebowitz, J. L., Tumulka, R., & Zanghì, N. (2020). Gibbs and Boltzmann entropy in classical and quantum mechanics. In *Statistical mechanics and scientific explanation: Determinism, indeterminism and laws of nature* (pp. 519-581). Singapore: World Scientific.

Hausman, D. M. (1998). *Causal asymmetries*. Cambridge: Cambridge University Press.

Hausman, D. M. (2002). Review of *Physical Causation*. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 33(4), 717-724.

Hausman, D. M. & Woodward, J. (1999). Independence, invariance and the causal Markov condition. *The British Journal for the Philosophy of Science*, 50(4), 521-583.

Hausman, D. M. & Woodward, J. (2004). Modularity and the causal Markov condition: a restatement. *The British Journal for the Philosophy of Science*, 55(1), 147-161.

Janzing, D., Chaves, R., & Schölkopf, B. (2016). Algorithmic independence of initial condition and dynamical law in thermodynamics and causal inference. *New Journal of Physics*, 18(9), 093052.

Malament, D. B. (2004). On the time reversal invariance of classical electromagnetic theory. *Studies In History and Philosophy of Science Part B: Studies In History and Philosophy of Modern Physics*, 35(2), 295-315.

Maudlin, T. (2007). *The Metaphysics within Physics*. Oxford: Oxford University Press.

Ney, A. (2009). Physical causation and difference-making. *The British Journal for the Philosophy of Science*, 60(4), 737-764.

Paul, L. & Hall, N. (2013). *Causation: A user's guide*. Oxford: Oxford University Press.

Pearl, J. (2009). *Causality: 2nd Edition*. Cambridge: Cambridge University Press.

Price, H. (2007). Causal perspectivalism. In H. Price & R. Corry (Eds.), *Causation, physics, and the constitution of reality: Russell's republic revisited*. Oxford: Oxford University Press.

Roberts, B. (2019). Time reversal. *philsci-archive preprint: 15033*.

Roberts, B. W. (2017). Three myths about time reversal in quantum theory. *Philosophy of Science*, 84(2), 315-334.

Russell, B. (1912). On the notion of cause. *Proceedings of the Aristotelian society*, 13, 1-26.

Salmon, W. (1984). *Scientific explanation and the causal structure of the world*. Princeton: Princeton University Press.

Salmon, W. (1994). Causality without counterfactuals. *Philosophy of Science*, 61(2), 297-312.

Sklar, L. (1993). *Physics and chance*. Cambridge: Cambridge University Press.

Smith, S. R. (2003). Are instantaneous velocities real and really instantaneous?: An argument for the affirmative. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, 34(2), 261-280.

Struyve, W. (2020). Time-reversal invariance and ontology. *philsci-archive preprint: 17682*.

Tooley, M. (1990). Causation: Reductionism versus realism. *Philosophy and Phenomenolog ical Research*, 50, 215-236.

Uffink, J. (2007). Compendium of the foundations of classical statistical physics. In J. Butterfield & J. Earman (Eds.), *Handbook for the philosophy of physics* (pp. 924-1074). Amsterdam: Elsevier.

Wallace, D. (2011). The logic of the past hypothesis. *http://philsci-archive.pitt.edu/8894/*.

Williams, P. (2022). Entropy, complexity, and causal asymmetry in quantum theories. *Foundations of Physics*.

Woodward, J. (2003). *Making things happen: A theory of causal explanation*. Oxford: Oxford University Press.

Woodward, J. (2014). A functional account of causation; or, a defense of the legitimacy of causal thinking by reference to the only standard that matters—usefulness (as opposed to metaphysics or agreement with intuitive judgment). *Philosophy of Science*, 81(5), 691-713.

Woodward, J. (2022). Flagpoles anyone? Causal and explanatory asymmetries. *THEORIA. An International Journal for Theory, History and Foundations of Science,* 37(1), 7-52 (https://doi.org/10.1387/theoria.21921).

Porter Williams is an Assistant Professor of Philosophy at the University of Southern California. His research interests are in the history and philosophy of physics, particularly quantum theories, and the general philosophy of science.

Address: Department of Philosophy, University of Southern California, Los Angeles, CA 90089, USA.
Email: porterwi@usc.edu
ORCID: 0000-0003-0242-4045