

Back to Reichenbach

Carlo Rovelli

*Aix-Marseille University, Université de Toulon, CPT-CNRS, Marseille, France,
Department of Philosophy and the Rotman Institute of Philosophy, Western University, London ON, Canada,
and Perimeter Institute, 31 Caroline Street N, Waterloo ON, Canada*

In his 1956 book ‘The direction of Time’, Hans Reichenbach offered a comprehensive analysis of the physical ground of the *direction of time*, the *notion of physical cause*, and the relation between the two. I review its conclusions and argue that at the light of recent advances Reichenbach analysis provides the best account of the physical underpinning of these notions. I integrate recent results in cosmology, and relative to the physical underpinning of records and agency into Reichenbach’s account, and discuss which questions it leaves open.

I. INTRODUCTION

I address two questions in this article. The first is *in what sense and why is the future different from the past*. The second is *why a physical cause comes earlier in time than its effect*. I consider these questions from the perspective of physics. That is, I ask what is the physical basis of the manifest difference between the two directions of time, and what is the physical reasons why we say that a cause precedes its effect. More precisely: how these two factual aspects of our world (future and past differ, causes precede effects) follow from the laws of physics that we know and the contingent state of the world in which we happen to find ourselves. Strictly dependent on these questions are the physical underpinnings of notions such as record, memory, agency, and common cause.

A comprehensive discussion of these questions was given by Hans Reichenbach in ‘The direction of Time’ [1]. I briefly summarize Reichenbach’s conclusions and argue that they are convincing. They are solid on physical ground, and, as far as I can see, conceptually cogent. Perhaps because of Reichenbach’s meticulous style and of some a bit cumbersome technical constructions, Reichenbach’s account is not popular today, and not even much known. This is a pity, because it offers a clear solution to a number of issues currently rather confusingly discussed.

I integrate recent developments into Reichenbach’s account. I connect his key idea of the ‘branching’ structure of the universe to the timescale of the dynamics of the subsystems [2]. I place his thermodynamic picture into contemporary cosmology, showing that the thermal history of the universe vindicate and dramatically simplify Reichenbach’s assumptions [3]. I improve on Reichenbach’s discussion of agency by tying it to his discussion of records [4]. I also give a general discussion on why Reichenbach’s empirical approach to the direction of time and to the notion of cause is compelling and I ask which questions are left open by this account [5].

I give a short account of the ideas in ‘The Direction of Time’ (hereafter: DoT) below. This does not exhaust at all the comprehensive and acute discussion in the book, to which I urge the readers to refer. It only summarizes its key conclusions, which I believe are too often neglected

in today’s discussion on these topics.

II. BRANCHING STRUCTURE AND ENTROPY GRADIENT

No phenomenon reveals any detectable difference between the past and the future directions of time, unless it includes a process irreversible in a (general) thermodynamic sense. This is a fact. Here, ‘general thermodynamic sense’ means related to the existence of a large number of degrees of freedom treated collectively. It follows that when we talk about the direction of time we are talking about irreversible thermodynamic processes or by the direction of time indirectly defined by such processes. This is the fundamental observation¹ that renders Reichenbach’s analysis solid.

Thermodynamic irreversibility is described by the local increase of entropy in approximately isolated systems of physical variables. Hence, on empirical grounds, the very meaning of ‘future time direction’ can solely be ‘the direction in which the entropy of (most) isolated systems grows’ [DoT, p127].

Now, it is a fact of our world —discussed below— that entropy grows in the same direction in all sufficiently large isolated systems we encounter.² This fact equips the full temporal structure of our world with a single preferred direction. We then use this direction for describing reversible phenomena too. In short, on empirical grounds time direction is a thermodynamic phenomenon.³

The observation remains valid in quantum physics. Quantum dynamics does not distinguishes the past from

¹ See ample discussions in [6], [7] and [8].

² This makes sense in the Newtonian approximation, where a time variable can be globally defined. It also makes sense in special and general relativistic physics, since —at least in the currently accessible spacetime region— the pseudo-riemannian metric is time orientable. This article does not considers the properties of nature beyond the approximation where quantum gravitational phenomena can be disregarded.

³ The idea that a preferred direction of time might be intrinsic to time itself, independently from the irreversible thermodynamic phenomena, and is only *revealed* by these phenomena, is discussed and criticised in Section IX.

the future [DoT, p208]. Quantum indeterminism does not pick a preferred time direction: predictions and post-dictions are equally underdetermined [9]. The formalism is often presented in terms of *pre*-dictions only, but this is only because we tend to consider past events to be ‘fixed’, and future events to be ‘open’ [10], a distinction itself rooted into thermodynamics, as discussed below.

Thermodynamic phenomena are statistical in nature. Therefore the direction of time is only manifest statistically. Because of the large numbers due to the smallness of microphysics with respect to our scales, this statistical nature is largely hidden beyond the cogency of the second law of thermodynamics. But this does not question the statistical nature of the direction of time: it only hides it, rendering it counterintuitive.

The universe around us contains a large numbers of subsystems that are approximately isolated during certain time intervals (from a glass of water with ice cubes, to a galaxy). In general these systems have been in interaction with the rest of the universe in the past. Importantly, many of these systems, are not at, or near, thermodynamic equilibrium. (Neither the water with ice cubes, nor a galaxy, are.) This is because we observe them at time scales that are shorter than their thermalization (or relaxation) time τ . The sun is still far from thermodynamic equilibrium after billions of years of existence. Any large portion of the universe can be considered approximately isolated: it is far from equilibrium, as it is full of burning stars, slowly converting hydrogen into higher-entropy elements. All these systems are therefore on a slope of their entropy curve.⁴

Reichenbach denotes these feature of the universe we inhabit as a ‘branch system’. More precisely a branch system is a physical system which [DoT, p136]: (i) is evolving along a slope of its overall entropy curve, (ii) it contains a large number of subsystems that ‘branch off’ for some long time intervals, namely they become approximately isolated during these intervals, and are such that (iii) time averages along their path are appropriately captured by averages over the systems themselves and (iv) their entropy is low at one branching point. Assumption (iii) is a kind of ergodic hypothesis to which Reichenbach devotes much technical details in the book. It is what allows us to determine statistical properties as averages over systems. Reichenbach shows that these conditions imply that in the vast majority of branch systems the direction in which entropy increases is the same, thus defining a common direction of time, determined by the entropy slope of the overall system.

This branch structure is realized in the universe around us, thanks in particular to its dynamics. For an isolated system S , the relaxation time τ is determined by the system’s hamiltonian. For instance the diffusion equation in

a gas depends on the microphysics and, in turns, determines the mixing time at macroscopic scales, hence the relaxation time τ of a mixture. Similarly, the interactions between a system and its environment determine the time τ the system takes to thermalize with its environment.

Consider an object at temperature T , separated from an environment at a different temperature T_{en} by a divider with small but non-zero thermal conductivity. At time scales t such that

$$\tau < t < \tau \quad (1)$$

the system can be considered as having a temperature $T(t)$ that evolves as

$$T(t) = T_{en} + (T(0) - T_{en}) e^{-t/\tau}. \quad (2)$$

At these time-scales, a direction of time is determined by the irreversible decrease of $\Delta T(t) = |T(t) - T_{en}|$. The direction of time determined by this irreversible phenomenon is the one for which

$$\frac{d}{dt} \Delta T(t) < 0. \quad (3)$$

Notice that at time scales much shorter than τ the temperature of the system is not well defined, while at time scales much larger than τ the system is in equilibrium with the environment and there is no change in macroscopic variables: nothing defines a direction of time anymore. A direction of time can be defined empirically by phenomena at these intermediate scales [DoT, p125]. There is no detectable preferred direction of time in a system fluctuating around equilibrium [DoT, p129], nor in a purely mechanical system. Long thermalization times yield regimes where time has a direction.

The point has been repeatedly emphasized by Eddington [11], Feynman [12], and many others: there is no clock without thermodynamic irreversibility. The mechanism that counts the oscillations of the oscillator of a clock, for example, could not work without dissipating energy.⁵

Realistic values of the quantities τ and τ in the universe range from nanoseconds to billions of years and more, making the arrow of time very well defined at our scales.

Notice that the slow equilibration implied by long thermalization times determines phenomena that strictly speaking are outside the regime of equilibrium thermodynamics, but without the complexities that the expression ‘non-equilibrium thermodynamics’ commonly suggests. There is nothing unclear in a regime as the one described by equation (2), for instance. The phenomena in these regimes are the phenomena most relevant for understanding the physics of the arrow of time.

⁴ Given macroscopic variables A_n , the entropy of a system is defined also for a microstate s away from equilibrium. This can be done in terms of the phase-space volume where $A_n = A_n(s)$.

⁵ Dissipation is used in the conversion of oscillatory (non-oriented) motion into monotonic (oriented) motion. Hence the *duration* of a time interval is determined by the mechanical device, but the *direction* of the passage of time is measured by the dissipation.

Importantly, when $t \ll \tau$ a system can find itself in a metastable state. This is a ubiquitous occurrence: almost all systems that we call ‘at equilibrium’ are really in metastable states: the reason they do not jump to higher entropy configurations is just because potential barriers make the jump improbable on relevant time scales. Environmental changes can drastically modify τ , throwing the system into a higher entropy state: a match can burn a pile of wood that has been resting for decades: the wood was in a metastable state, not in an equilibrium state.⁶ The existence of long thermalization times is the essential and commonly neglected ingredient of the temporal features of our universe.

In summary, a direction of time is only manifest in thermodynamic processes. It is determined by the common versus of the entropy slope in the subsystems forming the branch structure of the universe. Ubiquitous long thermalization and relaxation times keep subsystems isolated and far from equilibrium, and determine steady states regimes that are time oriented.

III. COSMOLOGY: THE ROLE OF THE SCALE FACTOR

The basic features of the thermal history of the universe revealed by modern cosmology are relatively simple, and nicely underpin Reichenbach’s account.

The history of the universe can be approximated by a model defined by an expanding homogenous space containing matter that interacts via the particle-physics standard-model interactions, and via the newtonian interaction. The expansion can be described by the metric

$$ds^2 = -dt^2 + a^2(t) d\vec{x}^2, \quad (4)$$

where the ‘scale factor’ $a(t)$ varies in the proper time t .

The other degrees of freedom of gravity and its strong-field features, manifest for instance in gravitational waves and black holes, do not appear to have played any essential role in the thermal history of the universe so far. This fact, by the way, casts doubts on the attempts to trace the observed arrow of time to them.⁷

⁶ Some authors appropriately distinguish the second principle of thermodynamics, which states whether or not a thermodynamic process is possible, from an ‘Equilibrium Principle’ [13] or ‘Equilibration Principle’ [14], which states that systems spontaneously attain their equilibrium state. Strictly speaking, this is not what we see around us: systems do spontaneously move towards higher entropy states, but they mostly get trapped into metastable states, during the time scale we observe them.

⁷ An example of these attempts is the idea that it is the smoothness of the geometry that provides the low entropy of the universe relevant to account for the observed arrow of time [15]. The problem with this idea is that if this was the case, a universe where the only dynamical part of the gravitational field was $a(t)$, and matter interacted gravitationally via a Newtonian interaction only, would have no arrow of time, contrary to what the analysis of this model shows.

According to the current understanding of cosmology, matter was at thermal equilibrium in the early universe. This may seem paradoxical, given the fact that entropy had to be low in the past; but the contradiction is only apparent. The reason is entirely in the dynamics of the single degree of freedom $a(t)$. Homogeneity implies that a co-movig region R can be considered isolated. The volume of R scales as $a(t)^3$ and $a(t)$ has grown much faster than the thermalization time scale

$$\frac{da}{dt} \gg \frac{a}{\tau} \quad (5)$$

during the early cosmological expansion.⁸ This is an irreversible process. The air rapidly expanded and compressed while inflating a bike’s tire with a hand pump warms up: there is dissipation, which signals irreversibility. The rapid expansion of $a(t)$ throws matter out of equilibrium [3, 16].

This has happened in a particularly consequential manner during the nucleosynthesis: the expansion has been too rapid for the ratio between hydrogen and helium densities to reach its maximum entropy value. The result is that there is far more hydrogen than what equilibrium would require: matter has been thrown out of equilibrium during the nucleosynthesis. With the expansion (and the consequent decrease in temperature), the relaxation time $\tau_{H/He}$ to the hydrogen↔helium thermodynamic equilibrium is huge —much larger than cosmological times—, bringing and freezing the universe into a metastable state.

Later, however, Newtonian gravitational instability, coupled to dissipation due to radiation, compresses matter and raises its temperature inhomogeneously. At higher temperature $\tau_{H/He}$ drops drastically (a dynamical effect) and the $H \rightarrow He$ process fires up, further increasing temperature, self sustaining, and rapidly increasing entropy. This process is of course called a ‘star’.

In other words, stars are the regions where the large amounts of free energy that the rapid initial expansion has frozen into the τ_{H-He} thermodynamical imbalance during the nucleosynthesis, get liberated, fuelling strong irreversible processes. The free energy liberated by the Sun fuels the entire thermodynamics of the biosphere, to which we belong. The irreversible processes that make us are therefore ultimately fuelled by the early-universe smallness of $a(t)$, via the nucleosynthesis and the burning Sun.

The scale factor $a(t)$, however, is not an *external* parameter in the cosmological dynamics. It is a dynamical variable coupled to the matter degrees of freedom by the Friedman equation (hence ultimately by the Einstein equations, of which the Friedman equation is a special case). It is therefore proper to consider the matter degrees of freedom and the scale factor as components of a

⁸ I am not referring to inflation here.

single interacting thermodynamical system. In the early universe this system was *not* at equilibrium. It was very far from equilibrium, because the scale factor degree of freedom was badly out of equilibrium with the matter. This is where past low-entropy dwells.

An illuminating analogy is given by a gas in a volume, closed by a piston attached to a spring. Say at some time the gas is in equilibrium with itself but the spring has far more (kinetic and potential) energy than what thermal equilibrium would demand. Then the gas is rapidly thrown out of its equilibrium by the rapid motion of the piston. This is what has happened to the matter of the universe.

The analogy fails in two respects. First, the sign of the energy transfer is reversed: in cosmology matter loses energy in its interaction with the scale factor, and cools. Second, the cosmological system has actually no equilibrium state.⁹

Thus, a co-moving region of the universe: (i) was overall in a low entropy state in the past, *just because of the smallest of $a(t)$* ; (ii) it is, since, on an increasing entropy slope; (iii) the fast volume expansion sent the matter out of equilibrium even if matter was by itself in equilibrium to start with; (iv) the initial out-of-equilibrium condition provides the free energy that gets transmitted to the matter by the rapid expansion and frozen by relaxation times much longer than the current cosmological scale; (v) newtonian gravitational instability generates subsystems approximately isolated, and triggers irreversible phenomena that liberate free energy, which in turns fuel irreversible physics.

Importantly, these are features of generic dynamical histories of the universe only limited by having a large variation in the value of $a(t)$ during the time interval considered.

Now, these features realize Reichenbach's branch system. They show that a Reichenbach's branch system is implied by a cosmology that is defined by the dynamical laws we know and a history that is rather generic, except for the fact that at some time the scale factor was much smaller than now. Cosmology underpins Reichenbach's assumptions.

The role of cosmology in discussing the arrow of time, on the other hand, should not be overemphasized. The direction of time is a phenomenon that is observed locally. We see irreversible processes around us, and we see all of them consistently oriented. We see that the universe is formed by approximately isolated subsystems, out of equilibrium with respect to one another, with long thermalization times and a common orientation of their entropy gradient. The fact that this phenomenology is nicely accounted for by current cosmology, and can be resumed in the information of the small initial value of $a(t)$, adds interesting information about it, but is not

needed to capture what we mean *locally* by the arrow of time.

In fact, restricting the gravitational field dynamics to the sole scale factor is of course an approximation. It suffices to account for what we see, but it misses important aspects of nature, two in particular: those captured by special relativity and those captured by general relativity. Both are relevant, and both show that the picture of a single time or preferred time variable is only an approximation. Cosmological time is an approximate notion making sense only within the rough homogeneity approximation: it is the age from the big bang, but when two galaxies meet (as the Milky Way and Andromeda are heading to), their age from the big bang is in general different.

Because of special relativity, there is no common present, and therefore no common past. The notion of past is not only relative to a time but also to a spatial location. As noticed in [DoT, p85] and discussed in detail in [18] this implies that at any point p of a spacetime the future is strictly unpredictable, because the past of any point p' in the future of p is larger than the past of p , and therefore the information about the past available (thanks to records of the past, see below) at p is always insufficient to compute what happens at p' (which is dynamically determined by the past or p' , not p). No Laplace demon located in spacetime can compute its future.

General relativity questions the picture of a global time evolution even more radically, because in general there is no global proper time that can play the role of the independent time t of the cosmological picture given above.

And finally quantum mechanics introduces an element of irreducible indeterminism that defeats any hope to see the present as already determined by some past.

All these physical facts force us to think about physics locally, rather than globally. Locally, the direction of time is determined by the coherence of the entropy gradients we observe. Cosmology tells us that, given what we know of the universe, this coherence is unavoidable if in the past the scale factor was much smaller than today.

IV. RECORDS

A branch system with weakly interacting branches and long relaxation times is a sufficient condition for the existence of abundant records, or 'traces', of the past. In other words, records of the past appear naturally in a branching system, on time scales smaller than τ and τ . The mechanism that brings this about is sketched in [DoT, p152 and p179] and quantitatively detailed in [2].

A record of the past in an 'improbable' configuration (a photo, a text, a step in the sand, a crater on the moon), hence out of equilibrium, which in the past branched off from the environment (or was affected by the environment) and where a correlation between the branching

⁹ This is a fact which ultimately might be relevant [17], but not at this stage of the discussion. I will return to this point later on.

event and the configuration of the system perdures, protected from dissipation by a long τ , and a long τ . The improbability of the record stems from the low entropy of the branching point. See [2] for an explicit model illustrating how records happen naturally in a branch system.

A record carries information, in Shannon's sense of *relative* information (that is, correlation). The information coded in the record is paid for by the increase in entropy at the moment of the formation of the record. The amount of information I in a record is therefore bounded by thermodynamical variables. If the memory system has temperature T_m and the environment has higher temperature T_e , it is shown in [2] that the information that can be stored in the memory system is bounded as

$$I < \frac{C_m(T_e - T_m)^2}{kT_e T_m} (1 - e^{-\tau/\tau}). \quad (6)$$

where C_m is the heat capacity of the memory system and k the Boltzmann constant. Records are therefore natural mechanisms converting past low entropy into macroscopic information.

In fact, everything we consider information is embodied in records, including books, brains, memories, culture, music, or DNA, and is therefore all sourced from past low entropy.

A beautiful section of DoT is the detailed analysis of measurement instruments [DoT, p178]. Any measurement instrument *records* the measured quantity, and therefore is an example of formation of records. Hence any measurement instrument works thanks to thermodynamics. This can be seen by imagining of running a measurement process backward in time: we obtain a process that is dynamically possible (dynamics is time reversal invariant) but highly implausible, because of the improbability of the strange coincidences in the evolution of the instrument and the system converging: an implausible lowering of entropy.

A consequence of the existence of records, is that they make us aware that entropy was lower in the past. This is a deduction that would be impossible in the absence of a branch structure [DoT, p129]. Imagine a cup of tea left for a long time in an empty room. The tea evaporates, raising the entropy. From the final configuration of the water vapour mixed with air, there is no way to deduce backwards the initial lower entropy situation. But the tea may have left traces on the cup, from which we can infer that the cap was full of tea sometimes in the past. So we infer a lower entropy state from a trace. This is how we know that the universe had low entropy in the past: astronomical observations carry information about the past universe to us. Light from a distant galaxy, in this perspective, is a branched system that has not thermalized with the rest of the universe, and is therefore bringing us direct information about the past. If the traveling photons thermalized with the rest of the universe, we would miss this evidence.

Memory of course is a specific example of records. Hence memory is permitted by the entropy gradient.

This is why we have no memory of the future. Memory gives us a relatively clear picture of the past (and not the future), and from this picture we get the feeling that the past is 'fixed' (while the future is not). More on this later.

V. CAUSES AND ENTROPY

Chapter IV of DoT asks what we mean when say that an event A is the *cause* of an event B. This is of course a vague question given the multitude of different meanings that we attribute to the noun 'cause' in common language. The classification of the diverse distinct uses of this noun is an exercise for philosophies ranging from Aristotle to Buddhism. Here I restrict to the numerous instances where A is a physical cause seen as necessarily earlier than its effect B.

Famously, Hume considered the possibility that by causation we only mean correlation, and the distinction between the cause and the effect is just verbal: we *call* cause the correlated event happening earlier in time [19]. The view has been revived by Russell, on the basis of the idea that in physics causation is only captured by laws which express nothing else than correlations between events [20]. Hume and Russell have a point, but miss something: their view has been correctly criticized, by pointing out that when we talk about causation we mean more than correlation (for instance: [21]). Indeed, not only we distinguish sharply between causation and correlation in many sciences, but that distinction is often precisely what is of interest: does smoking *causes* cancer, or is there just a correlation between cancer and smoking, without direct causation? Figuring out that the first option is the right answer has saved very many lives. Hence causation is more than just correlation.

A convincing and scientifically fruitful investigation of the notion of causation has been developed in recent years by modelling networks of probabilistically correlated events and using the notion of *intervention* [22–24]. An event A is understood to be a cause of an event B when the correlations are such that if I forcefully change A, then B changes. Stopping smoking does decrease the probability of dying by cancer. It wouldn't if this correlation was not causation. These ideas (already clearly expressed in ToD, p43, 197, 204-205) are clarifying, but they rely on one assumption: that interventions affect the future and leave the past unchanged, which is taken as an assumption in all these modelings of causal networks. Hence these models do not explain why causes precede their effects: they assume so.

So, why do interventions change the future and leave the past unchanged? To answer this question requires to understand what an intervention is, on physical grounds. Two side observations are important before answering this question.

The first is that although perhaps anthropomorphic in its origin, the notion of intervention does not require any

actual anthropomorphism. If I study the geology of the moon, I can consider the fall of a meteorite as an intervention. I can ask what would change if the meteorite had or had not fallen, and I can study the effects *caused* by the fall of the meteorite. The point is here a (arbitrary) split of the world into a part under study whose regularities we follow (the geology of the moon), and a part *considered* external and accidental (the meteorite) because we are not interested in (or we do not have sufficient data about) it. The notion of intervention depends on this (arbitrary) split. Nothing anthropomorphic.

The second observation regards counterfactuals. The language of counterfactuals (for instance understood in terms of possible worlds [25]) may be convenient, but is not needed. We make statements about causation because we control world's *regularities*, and these have been deduced via observations, observed frequencies (Reichenbach is a frequentist about probability) and induction. It seems to me that we can translate counterfactuals into statements about regularities, extrapolations and maybe subjective expectations motivated by these.

With these preliminaries set aside, I return to the question: why do interventions change the future and leave the past unchanged?

Trying to find the answer in subjective perspectivalism might be tempting but is wrong. The fact that interventions affect the future and not the past is an objective feature of the world around us. The interaction of a stone (intervention) with a pond of water is correlated with concentric waves *in the future*, not in the past, of the interaction.

But the causal nexus cannot be purely mechanical, because, again, mechanics is time reversal invariant. Hence, it can only be, once again, thermodynamical, hence statistical.

First, before receiving the stone, the pond is *not* in equilibrium with its environment, which includes the stone. If it was, there would be no way to distinguish wave configurations before of after the fall, because the interaction with the stone would be just a generic thermal fluctuation: the stone could be emitted by the pond with the same probability it is received. If a gas at temperature T is hit by a molecule coming from a gas at the same temperature T , no concentric waves form, because the effect is fully hidden by the thermal fluctuations in a way that does not permit us to distinguish the future from the past. It is the thermodynamic unbalance between the pond and the kinetic energy of the stone that makes the time direction detectable. The thermalization time τ is much longer than the time scale of the observation.

Second, the train of expanding waves is itself not at equilibrium. When reaching equilibrium, its energy is dissipated into pond and there isn't any detectable effect of the past interaction anymore. The train of expanding waves is a subsystem whose relaxation times τ and τ are longer than observation time. It is a branched system in the sense of Reichenbach, that detached from the rest

(for dynamical reasons) at the moment of the interaction with the stone, in a low entropy context (pond and stone out of equilibrium).

The example indicates the general characterisation of physical causation: The cause is the interaction at the low entropy end of a branch of an isolated system that find itself on a low entropy (an ordered) configuration, which is the effect. This is the definition of cause in [ToD p151], and is the physically correct one: when we talk of physical causes producing a physical effect in the future, we are talking about that. All physical 'effects' that we recognize are ordered (low probability) states of affairs in systems that maintain the imprint due to an event in a past lower entropy configuration of the world.

The similarity of the notion of record with the notion of physical cause is now obvious: the effect is a record of the cause. The recorded event happens to cause the record.

But this is not the full story. There is something else, which is the subtlety that creates much confusion. We *also* extend the notion of cause to indicate correlations in *reversible* processes. For instance, in the collision of two elastic balls we say that the collision has caused one ball to take a new direction *in the future*.

Now, *this* use of 'cause' is perspectival, and in *this* case Hume and Russell are literally right: the distinction between cause and effect is only terminological. We call 'cause' the term of the correlation that happens first, and nothing else that the time ordering distinguishes cause from effect. The time ordering we use, in these cases, is the one that is defined by all the irreversible processes, and which comes natural for us to think in terms of (see later on on this). So, in this context, as in all contexts where there is no dissipation, the time arrow and the arrow of causation are just perspectival: we simply *interpret* phenomena in a certain time direction. When we employ an *oriented* notion of causation to phenomena where dissipation is negligible we are simply 'psychologically speaking' transferring the orientation of time determined by dissipation to these phenomena [ToD, p156].

But careful: thinking that this is always the case fails to capture the fundamental fact that there *is* a preferred time direction in the nature around us, which depend on the entropy gradient, and this determines a preferred direction in the formation of records and in the observable phenomena correlated to interventions. This is what grounds the proper notion of causation that we employ in science and in everyday life (for a good discussion on the distinction between these two distinct cases, see [26].)

The bottom line is that time-oriented causation is a thermodynamic notion, rooted in the thermodynamic structure of the world, and in particular its entropy gradient.

The direction of time is tied to causation because both are determined by entropy gradient of our world and by its branch structure, which is due to the actual dynamical laws of the world.

VI. COMMON CAUSE IS NOT AN ASSUMPTION, IS A THEOREM

There is one detail in Reichenbach analysis of causation that has become widely known: his notion of common cause. Reichenbach enounces it as a principle in ToD [p157]. When the joint probability $P(A, B)$ of two dynamically independent events A and B is higher than the product of the probabilities of the two

$$P(A, B) > P(A)P(B), \quad (7)$$

we search for the reason of this unlikely correlation in a *common cause*, and we always search for this common cause *in the past*. If two students turn back the final exam with precisely the same mistake and the same wording, this looks improbable, because given the low probability of each specific wording, the sameness should be very improbable. We suspect something has happened *in the past* that could ‘explain’ the coincidence. This is not a perfect proof of cheating (coincidences happen — causation is statistics), but it is a strong element of evidence.

In the way the Reichenbach’s common cause principle is often mentioned today, the fact that the common cause is *in the past* is taken to be a self standing principle. Nothing of the sort in Reichenbach. Let me quote him for once

‘This principle does not represent a new assumption, but is derivable from the second law of thermodynamics, if this law is supplemented by the hypothesis of the branch structure.’ [DoT, p157.]

I refer the reader to ToD for the proof [DoT, p163], but the core of it should be clear. The coincidence requires something less probable, and this can be found in the past, because of the past low entropy. The branching structure provides the concrete mechanism that promotes thermodynamical low entropy into macroscopic information, hence into unlikely configurations. The common cause is in the past because in the past entropy was lower.

VII. EXPERIENTIAL TIME

There are two ways what has been said so far affect our experience of time. The first is that we live in a world full of irreversible phenomena, namely phenomena where the total entropy increases. Because of this fact, we consider, by habit, these phenomena plausible and their time reversal implausible. In other words, it seems natural to us that strange coincidences do not happen in the future, but are ok in the past. Why we consider this plausible? Because this is what de facto happens in a world where the scale factor has grown as it has in our.

But the way the thermodynamic arrow of time is rooted into our thinking is far deeper than this. The very working of our brain is based on the existence of records.

(‘It is not a human prerogative to define a flow of time; every registering instrument does the same’ DoT pg270[.]) One possible way of understanding our brain, currently under intense investigation, is as a device shaped by evolution to be effective in trying to anticipate the future on the basis of past memories [27]. This is possible because records of the past exist. And they do because of the thermodynamical structure of the world. Hence our own thinking is an expression of the thermodynamical organization of the world. The information our brain works with is directly sourced from past low entropy, as discussed above.

In a world in thermodynamic equilibrium, or where dissipation and an entropy gradient were absent, not only there would be no way for a time direction to be defined, but there would also be no way for a thinking brain to work, either. As thinking entities, we are a product of the entropy gradient and the branch structure of the world. This is why the orientation of time is so natural to our intuition.

VIII. AGENCY

The above analysis sheds light on the notion of agency, leading to some important consequences, in my opinion.

Agency is a concept that is increasingly employed in foundational contexts, including the foundations of quantum theory (for instance in q-bism), thermodynamics (in the definitions of macroscopic quantities as those on which we can act), science itself (in particular in any instrumentalist or operational framework). Agency is the property of an agent that is taken as acting on a physical system, affecting its *future* evolution, and capable of doing so in different manners. The notion of agency is analyzed in thermodynamic terms in [4], where its connection to the phenomenology of records is pointed out.

Agency (like intervention, to which it is strictly connected) requires a (arbitrary) split of the world into a part considered to be the physical system ‘acted upon by the agent’, and a part including the agent. The notion of agent is relative to this split. The hammer is an agent for the nail, my hand is an agent for the hammer, my brain is an agent for my hand and your hammer on my head is an agent for my brain.

Calling a system an ‘agent’ amounts to disregarding the dynamical chains in which the agent is part, and hence *considering* it ‘free’. Such ‘freedom’ is the name we give to the fact that we disregard the mechanics, statistical, or other components of those dynamical chains. Or the name we give to the rich multilevel external and internal complexity of what brings about the agents choices [28]. In other words, a system is an agent to the extent that we avoid (by choice or by necessity) to fully unravel why it behaves as it does.

Agency is therefore fully compatible with determinism. In a deterministic world, the possibility for an agent to have the same macroscopic past but different

macroscopic futures is provided by the fact that the same macroscopic past is compatible with different microscopic pasts that can evolve in different macroscopic futures [4, 29]. However, this is of little effect, given that the actual world (or at least the part accessible to us) happens not to be deterministic, thanks to quantum theory.¹⁰

When seen as an agent, a system has a thermodynamical behaviour similar to records [4]: it promotes low-entropy into macroscopic information. Its different (‘free’) choices generate information by necessarily increasing entropy. They transform negative entropy into information, like records do. The number N of alternatives an agent can choose from satisfies [4]

$$N < 2^{\frac{t}{\tau} \frac{T_a - T_{en}}{T_{en}}}, \quad (8)$$

where T_a and T_{en} are the temperatures of the agent and the environment, τ the thermalization time of the agent and t an average interval between agent-environment interactions. The generation of information in a choice, and the time orientation of agency (agency affects the future) necessarily require that irreversible phenomena are in play, hence an entropy gradient. The model in [4] shows that it is the entropy gradient and the brunch structure that permits agency. Agency is an intrinsically macroscopic notion.¹¹

The systems that we call agents have therefore an embodied time direction. Living agents may have memory, and a brain that dissipates energy. Agents are time-oriented by the ubiquity of irreversible phenomena in a number of ways: by their memory, by the thermal unbalance with the environment, and by the fact that their interventions affect the future, as detailed above.

I think that this observation is important for the following reason. If we use agents in the foundations of a discipline, we seriously risk to make the mistake of projecting the specific property of the agents to the entire domain of investigation.

This is a common mistake precisely for what concerns the direction of time. In doing science we treat ourselves as macroscopic time oriented free agents. Indeed we are so, by our memory and by the irreversible phenomena that make us. Hence we formulate the scientific inquire in time directed terms (we say that equations of motion or quantum transition amplitudes ‘predict the future’).

¹⁰ This remains true in an appropriate sense even in ‘deterministic’ interpretations such as Bohmians or Many Worlds, because the total state that evolve deterministically is *in principle* inaccessible to us. Hence our best predictions remain probabilistic even within these interpretations.

¹¹ But of course any agent is also a physical system like any other. As such, it does not necessarily need to be seen as a ‘free’ agent. A complex system with a rich internal dynamics can be considered as an agent, as we do with humans, and can be analysed at many different levels, in terms of the combination of factors determining its behaviour, including its mechanics, its memories, the external influence, and much more.

This leads us to the wrong conclusion that nature is universally time oriented, blinding us to the fact that what is time oriented is ourselves (see [26]).

But science at its best is precisely freeing ourselves from these mistakes, and recognizing the perspectival aspects of the phenomena. This is why I think using agency in the foundations is something to be handled with care. An operational or an instrumental foundation of physics projects onto nature a fundamental time orientation which is a property of us observers, not a property of nature in general.

IX. IS THE TIME ARROW PRIMARY?

This is the good moment to pause and address a philosophical question. Why not assuming that time is oriented by itself, and its orientation *determines* the phenomenology described? That is, why not saying that time passes from the past to the future (whether there were irreversible phenomena or not), that the past is intrinsically fixed and the future open, and the phenomenology described (the growth of entropy, the branch structure of the universe, all time oriented phenomena) is such just *because* of an intrinsic orientation of time?

After all, this is our intuition. Causes precede effects *because* times goes from past to future, entropy grows *because* time passes and hence things get disordered, and so on. Isn’t the a priori difference between past and future the *explanation* of irreversibility, rather than the other way around (as argued in DoT and here)? This seems not only an intuitive position, but also an economical one: one simple assumption (the past is different from the future because time flows towards the future) seems sufficient to deliver all the rest. Why not just saying this?

The answer is that saying so is –at a moment of reflection– meaningless. It is the same as Molière famous explication of the reason why certain herbs makes you sleepy: ‘because they have a dormitive virtue’. Which is to say that they do so because they do so. The passage of time is the name we give to a phenomenology. It makes no sense, then, to take this name as the explication of the phenomenology. If it is not just the name we give to the phenomenology of irreversibility, what does it mean that time itself is oriented? Nothing at all that I can understand.

It does makes sense to use the language ‘because time has passed’ to connect a particular phenomenon to its class. We say ‘the tea has evaporated because time has passed’. But this is not an explanation of why irreversible phenomena have a direction: it is the name we give to a general fact.

One of the reasons this is very important is in order of not to be misguided in our search to extend the physical laws we know. In particular, we know that the temporal structure that characterizes the solutions of the Einstein equations is an approximation. It is not anymore present, in general, when quantum gravitational phenomena can-

not be disregarded. In quantum gravity, in general, there is no background spacetime, nor a dynamical spacetime metric to which we can attribute a direction of time [30]. In this context even the linear structure of time must be extracted, in general only as approximation, from the phenomena, and cannot be presupposed. Being chained to the natural intuition that the direction of time is foundational aspect of nature blocks us from developing the right theory.

The intuition of the marching ahead of time in the universe is a powerful one. It is not an empty intuition, but what it refers to is just the coherent increase of entropy in virtually all the branches we access. Trying to disconnect an intrinsic property of time from this phenomenology is like trying to disconnect the movement of the sun in the sky from the rotation of the Earth. It may be counter-intuitive, but the rotation of the sky is *nothing else* that the effect of the rotation of the Earth on what we observe. It is counterintuitive, but the one directional passage of time is *nothing else* that a bunch of macroscopic features of a generic history of the universe with a much smaller $a(t)$ at some point.

The expression ‘now’ is indexical (Reichenbach calls it ‘token-reflexive’ [DoT p270]): its meaning depends on the context, in particular the spatiotemporal context, in which it is pronounced. The main source of confusion about time comes from imagining that there could be a meaning of ‘now’ outside this context. Any act of language or act of thought is embedded into nature and into its spatiotemporal structure. Forgetting this causes all sort of paradoxes (as those at work in [31]’s A series.)

It is tempting to consider our experience as universal. It is tempting to interpret all the features of experiential time as universal aspects of nature, but there is no reason I could see for which it should be right. Why should our instinctive conceptual structure reflect nature at large? We are part of nature, but a very limited part of it. When we talk about time and its direction we are talking about *some* aspects of the concrete phenomenology of nature. The alternative idea that we can *directly* mentally experience a structure of the world more profound than this phenomenology makes no sense. How could we?¹²

We should adapt our intuitions and our concepts to our scientific discoveries, and not try to force our scientific discoveries into our weak and often misleading intuition and our *a priori* concepts.¹³

X. IS THERE A MYSTERY IN THE FACT THAT ENTROPY WAS LOW IN THE PAST?

It is a fact that in the early universe entropy was lower than today. A generic world satisfying the dynamical laws of physics that we know today, and where the scale factor has scaled considerably, displays all thermodynamic features (the branch structure, the arrow of time, the records of the past and the time-oriented causation) that characterize our world. Is this a satisfactory understanding of the arrow of time and the arrow of causation? In this section I present a few open ended considerations in this regard.

There is tension between the ‘generic’ in the paragraph above, and the ‘smallness of scale factor’. If we count ‘being generic’ as a good explanation, then why the value of the scale factor is not such that the state was ‘generic’ in the early universe? This is the reformulation of the old question: why was the entropy of the universe low in the past?

Having understood that the smallest of the entropy of the early universe is entirely due to the smallness of a single dynamical variable does tame the mystery of the past low entropy. To be ‘low entropy’ is to be ‘special’. The smallness of early universe entropy compared to now is colossal, and this sometimes suggests that the early universe needed to be ‘extremely peculiar’. But it was only ‘peculiar’ in so far as it had a small volume. The number of possible configurations of a gas of N molecules and energy U in a small volume is enormously smaller than the number of possible configurations of the same gas and the same energy in a larger volume. (The entropy of an idea gas with N molecules increases by $Nk \ln 2$ by doubling the volume.) The ‘past hypothesis’ [7] of our universe does not require any strangely peculiar initial state: it can be simply stated saying that $a(t)$ was much smaller than today.

But notice also that the scale factor does not appear to have any equilibrium value anyway [17]. Hence the question ‘why are we not at equilibrium?’ may have the simplest of the answers: because there is *no* equilibrium, given the actual dynamics of the universe (which includes gravity). Namely there is no value of the macroscopic variables (including $a(t)$) that is a maximum for the entropy. The phase space, in other words, might be infinitely large.¹⁴

Furthermore, the right question —if there is any— is not ‘how special was the state of the early universe’. The right question is how special is the solution of the equations of motion on which we happen to be. It is not clear to me how to make sense of this question neither in a rigorous nor in an intuitive manner.¹⁵

¹² It is beautifully put in [26]: ‘this view [...] requires a deep link between the mental, on the one hand, and some deep and fundamental time-asymmetric aspect of physical reality, on the other—without the time-asymmetry concerned being manifest at intermediate levels!’

¹³ A point nicely argued for instance is [32].

¹⁴ The phase space is a modal notion, hence this is a potential, not an actual, infinity.

¹⁵ I return to Penrose suggestion that our universe is ‘special’ be-

There is one last consideration which seems crucial to me. The conceptual construction described above relies on thermodynamics, hence ultimately statistics, because thermodynamics is based on ignoring some degrees of freedom, or treating them collectively. Statistics is unavoidable in principle because of quantum theory and *de facto* even in classical theory. But the statistics that underpins thermodynamics is based on additional inputs with respect to dynamics. I am not referring to the probability measure one needs to postulate (on this, see the extensive discussion in [14]). I am referring to the distinction between accessible (‘macroscopic’) and inaccessible (‘microscopic’) variables. This distinction is not intrinsic to a dynamical system: it is over and above dynamics itself. What grounds it?

If a physical system is considered as formed by subsystems, then the distinction between accessible and inaccessible variables of a subsystem can be determined physically, as follows. Consider two systems S and O described by hamiltonian variables s_n and o_n respectively, and let the interaction Hamiltonian be $H(o_n, s_n)$. If the system O is in the configuration $\{o_n\}$, the effect of S on O is determined by single function $A(s_n) = H(o_n, s_n)$ and its gradient (which enters the equations of motion). Hence O probes a subset of variables of S only. As O moves across different values $\{o_n\}$, it may probe more variables. Let’s call these A_n . The A_n variables are the macroscopic variables *with respect to* O . That is, the only way other variables manifest themselves to O is indirectly, via the entropy $S(A_n)$ determined by the phase-space volume where the A_n variables have a given value. This is the way macroscopic variables are characterized in thermodynamics: they are the accessible ones. In fact, recalling the discussion about agency given above, we can consider O an agent by disregarding its dynamics, and then the A_n variables are precisely the handles the agent has to manipulate the thermodynamic, macroscopic state A_n of O .

These considerations make clear that the determination of the macroscopic variables, hence the very notion of entropy, can depend on the split S/O , and the entropy of S can depend on the system O with respect to which it is defined.

This is no surprise: the entropy of the air in a room defined with respect to an ‘observer’ O that interacts with the air via a piston that determines the volume, and a pressure sensor, is higher than the entropy defined with respect to a different ‘observer’ O' that in addition can measure the concentration of Oxygen in the air. Entropy is an objective quantity, but a relative one.

Now, the direction of time is a thermodynamic phenomenon, hence it pertains to macroscopic variables.

cause its geometry is smooth, while ‘most’ geometries are crumpled [15]. This is a brilliant suggestion, but –again– it does not seem to me to work. If *this* was the source of the arrow of time, then there would be no arrow of time in the approximation where the geometry remains smooth. And yet there is.

Therefore *a priori* the direction of time could depend on the system O with respect to which the entropy is defined. If ‘low entropy’ means to be ‘special’, what may be special is not the state of S , but rather the system O with respect to which entropy is defined, and the way it interacts with S . Every baby is very special for its mother. It is not the baby to special: it is one particular mother that is special with respect to that baby.

This is the possibility considered in [5]. Namely the possibility that the direction of time be ultimately perspectival. A grandiose phenomenon, but nevertheless a perspectival phenomenon, a bit like the rotation of the sky around us: grandiose, and yet just perspectival. How special –or how ‘natural’– is the coarse graining defining the entropy which happened to be low in the early universe and which is uniformly growing today? I refer the reader to the reference [5] for more details on this speculative idea, and do not discuss it further here.

XI. SUMMARY

This is brief recapitulation of the points discussed in the article.

- (i) The universe in which we live has a *branch structure*, with many approximately isolated subsystems that are not at equilibrium. They are on an entropy slope, all with the same sign.
- (ii) A generic evolution of a universe governed by the laws of physics we know has such a structure, provided that *the scale factor was much smaller than now* at some time.
- (iii) This branch structure is due to internal (τ) and external (τ) *long relaxation times*, with respect to the time scales of interest.
- (iv) The *direction of time* is a thermodynamic, hence statistical, phenomenon: it *is* the common sign of the derivative of the entropy slope, in all irreversible processes, in the subsystems we witness. The direction of time is nothing else than this phenomenology. Reifying it is useless and misleading.
- (v) The branch structure with its entropy gradient is sufficient to generate ubiquitous *records*. Records are out of equilibrium configurations, protected by long τ ’s, correlated with past interactions, and thus encoding information about that past.
- (vi) The *information* in records (and, in particular, memories) is sourced by the entropy increase at the moment of formation.
- (vii) *Agency* is defined by a (arbitrary) split between a system and an agent, with the two not in thermodynamic equilibrium. The effect of agency is in the future of its action, like records are, because of the entropy gradients.
- (viii) Physical *causation* is characterized as follows. An event A is understood to belong to the causes of an event B if an agent changing A changes B.

- (ix) *Physical causation is time oriented* because it relies on the time orientation of agency, which in turn depends on the entropy gradient. Hence time-oriented causation is ultimately a thermodynamic notion, rooted in the thermodynamic structure of the world.
- (x) A *common cause* that accounts for correlations between two events that are not directly causally connected can be found in the past (not in the future), because of the entropy gradient: the ‘improbability’ of the correlation needs to be sourced by the past low entropy ‘improbability’.
- (xi) *Experiential time* is strongly oriented not only because we witness ubiquitous irreversible phenomena in a markedly time oriented world, but more importantly because the very working of our brain depends on the entropy gradient and requires dissipation. The brain elaborates the information in memories and senses’ records, and these are all phenomena sourced by low entropy (by free energy). At equilibrium, we would not only fail to detect a time orientation in the phenomena: we couldn’t think.
- (xii) *The direction of time, physical causation, records, memory, agency, intervention, common causes, are all notion that pertain to macroscopical variables.* They have no significance for microscopical variables, unless we project *our* time orientation onto them.
- (xiii) *Entropy is to some extent perspectival* because it depends on a coarse graining, and a coarse graining is physically determined by the interactions with another system. The relevance of this fact for the above discussion is, in my opinion, not yet sufficiently explored.

-
- [1] Hans Reichenbach, *The Direction of Time* (Dover, New York, 1956).
- [2] Carlo Rovelli, ‘Memory and entropy,’ (2020), [arXiv:2003.06687](#).
- [3] C. Rovelli, ‘Where was past low-entropy?’ *Entropy* **21** (2019), [10.3390/e21050466](#).
- [4] Carlo Rovelli, ‘Agency in Physics,’ (2020), [arXiv:2007.05300](#).
- [5] Carlo Rovelli, ‘Is Time’s Arrow Perspectival?’ in *The Philosophy of Cosmology*, edited by Khalil Chamcham, Joseph Silk, John Barrow, and Simon Saunders (Cambridge University Press, 2016) pp. 285–296, [arXiv:1505.01125](#).
- [6] Huw Price, *Time’s Arrow and Archimedes’ Point* (Oxford University Press, 1996).
- [7] David Albert, *Time and Change* (Harvard University Press, Cambridge MA, 2000).
- [8] Carlo Rovelli, *The Order of Time* (Riverhead, New York, 2018).
- [9] Carlo Rovelli, ‘An Argument Against the Realistic Interpretation of the Wave Function,’ *Foundations of Physics* **46**, 1229–1237 (2016), [arXiv:1508.05533](#).
- [10] Andrea Di Biagio, Pietro Donà, and Carlo Rovelli, ‘The Arrow of Time in Operational Formulations of Quantum Theory,’ *Quantum* **5**, 520 (2021).
- [11] Arthur Eddington, *The Nature of the Physical World* (The University of Michigan Press, Michigan, 1928).
- [12] Richard Feynman, *The Character of Physical Law*. (BBC Publications, 1965).
- [13] Harvey R. Brown and Jos Uffink, ‘The Origins of Time-Asymmetry in Thermodynamics: The Minus First Law,’ *Stud. Hist. Phil. Mod. Phys.* **32**, 525–538 (2001).
- [14] Wayne C. Myrvold, *Beyond Chance and Credence* (Oxford University Press, New York, 2021).
- [15] Roger Penrose, ‘Singularities and Time-Asymmetry,’ in *General Relativity: An Einstein Centenary Survey*, edited by S W Hawking and W Israel (Cambridge University Press, 1979) pp. 581–638.
- [16] David Wallace, ‘Gravity, entropy, and cosmology: In search of clarity,’ *British Journal for the Philosophy of Science* **61**, 513–540 (2010), [arXiv:0907.0659](#).
- [17] Sean M. Carroll and Jennifer Chen, ‘Spontaneous Inflation and the Origin of the Arrow of Time,’ (2004), [arXiv:0410270 \[hep-th\]](#).
- [18] Jenann Ismael, ‘Determinism, Counterpredictive Devices, and the Impossibility of Laplacean Intelligences,’ *Monist* **102**, 478–498 (2019).
- [19] David Hume, ‘*Treatise of Human Nature*,’ (1736).
- [20] Bertrand Russell, ‘On the Notion of Cause,’ *Proceedings of the Aristotelian Society* **13**, 1–26 (1913).
- [21] Nancy Cartwright, *Hunting Causes and Using Them: Approaches in Philosophy and Economics* (Cambridge University Press, Cambridge, 2007).
- [22] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction and Search* (Springer Verlag, Ne, 1993).
- [23] Judea Pearl, *Causality: Models, reasoning and Inference* (Cambridge University Press, 2000).
- [24] J Woodward, ‘Causation and Manipulability,’ (2001).
- [25] David K. Lewis, *On the Plurality of Worlds* (Blackwell Pub, 2001).
- [26] Huw Price, ‘Causal Perspectivalism,’ in *Causation, Physics, and the Constitution of Reality: Russell’s Republic Revisited* (2007) pp. 250–292.
- [27] Dean. Buonomano, *Your Brain Is A Time Machine: The Neuroscience and Physics of Time* (W W Norton and Co Inc, 2017) p. 293.
- [28] Jenann Ismael, *How Physics Makes Us Free* (Oxford University Press, 2016).
- [29] Barry Loewer, ‘The Consequence Argument Meets the Mentaculus,’ (2020).
- [30] Carlo Rovelli, *Quantum Gravity* (Cambridge University Press, 2004).
- [31] John McTaggart, ‘The Unreality of Time,’ *Mind* **17**, 457–474 (1908).
- [32] Howard Stein, ‘Physics and Philosophy Meet: the Strange Case of Poincaré,’ *Foundations of Physics* **51** (2021), [10.1007/s10701-021-00460-x](#).