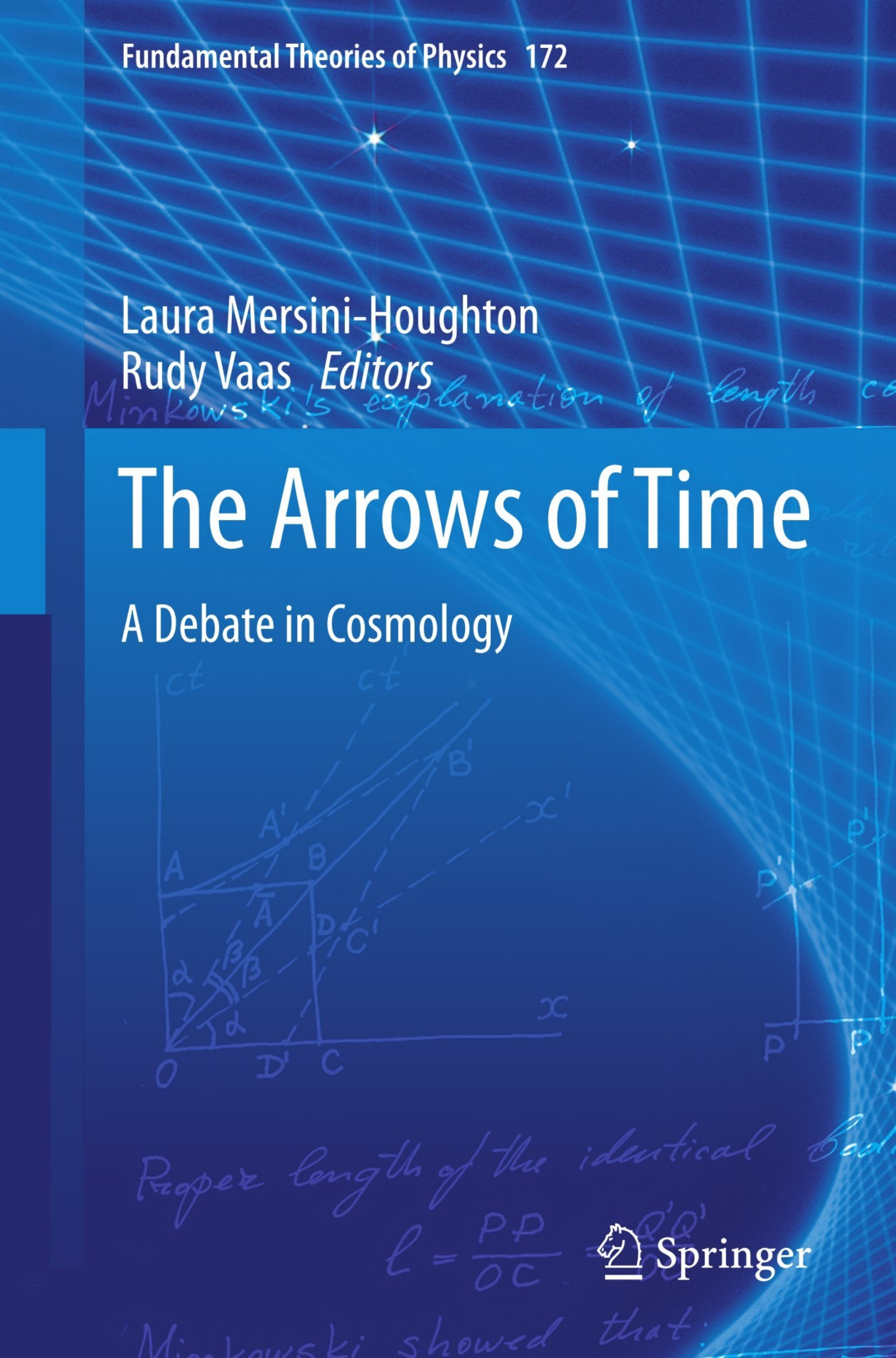

Fundamental Theories of Physics 172

Laura Mersini-Houghton
Rudy Vaas *Editors*

The Arrows of Time

A Debate in Cosmology

Springer

Fundamental Theories of Physics

The international monograph series "Fundamental Theories of Physics" started with a new Editorial Board from volume number 172 on. Earlier volumes which appeared in the names of the new Editorial Board were still reviewed and published under the responsibility of the previous Editors of the series.

Volume 172

For further volumes:
http://www.springer.com/series/6001

The series “Fundamental Theories of Physics” aims at stretching the boundaries of mainstream physics by clarifying and developing the theoretical and conceptual framework of physics and by applying it to a wide range of interdisciplinary scientific fields. Original contributions in well-established fields such as Quantum Physics, Relativity Theory, Cosmology, Quantum Field Theory, Statistical Mechanics and Nonlinear Dynamics are welcome. The series also gives a forum to non-conventional approaches to these fields. Publications should provide perspectives and carefully bridge conventional views with the presented new and promising ideas.

Although the aim of this series is to go beyond mainstream physics, a high profile and open-minded Editorial Board will carefully select the contributions and will ensure the high scientific standard of this series.

Laura Mersini-Houghton • Rudy Vaas
Editors

The Arrows of Time

A Debate in Cosmology

Springer

Editors
Prof. Laura Mersini-Houghton
University of North Carolina
Dept. Physics and Astronomy
Chapel Hill North Carolina
USA
mersini@physics.unc.edu

Rüdiger Vaas
bild der wissenschaft
Ernst-Mey-Str. 8
D-70771 Leinfelden-Echt.
Germany
Ruediger.Vaas@t-online.de

ISBN 978-3-642-23258-9 e-ISBN 978-3-642-23259-6
DOI 10.1007/978-3-642-23259-6
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2012939849

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Contents

Introduction

Time is of central importance to science and philosophy. And yet, the simplest questions – "is time real, or is it an essential part of the structure of human intellect?" – remain largely controversial. Theories of nature can be broadly categorized into two sets of information: physical laws which relate the sequence of states of a system, and initial conditions determined at a fixed moment in time. Clearly, a description of the succession of states or the choice of an initial moment where data about the system is defined, also involve time. Understanding this turns out to be as difficult as probing the origins of the universe since both physical laws and initial conditions assume a concept of time, which, in most cases, is inseparably interwoven into the theory and its predicted outcomes. Though time is ubiquitous and intuitive, it still defies comprehension. Disentangling ourselves from time to enable objective and independent investigation is the challenge.

Understanding the nature and the direction of time has occupied the minds of philosophers and scientists throughout history. It continues to do so. As far back as the fifth century, St Augustine wrote in *Confessions*, Book 11: "what then is time? If no one asks me I know. If I wish to explain it to one who asks, I know not… My soul yearns to know this most entangled enigma". The enigma persists. The yearning for understanding has now fallen on physicists working on the most fundamental questions about the cosmos and the origins of the universe.

What is the enigma of time?

1. The nature of time: is time an inherent and intrinsic ingredient of nature or did it emerge only at the big bang?
2. The arrow of time: why is there a clear direction from past to future, i.e. what breaks the time translation symmetry and sets an arrow of time? Why should the birth of the universe be determined by this direction?
3. The time-symmetry of physical laws: why is it that the laws of physics, which describe the universe we live in, cannot distinguish between past and future? How can physical laws respect time translation symmetry when the universe breaks it at the big bang?

L. Mersini-Houghton and R. Vaas (eds.), *The Arrows of Time*, Fundamental Theories of Physics 172, DOI 10.1007/978-3-642-23259-6_1,

This book describes contemporary views of physicists on the nature of time, the origin of time translation symmetry breaking, and the implications this enigma has on the origin and predictions of physical laws.

A consistent treatment of the three time enigmas described above, is presented in Laura Mersini's chapter in the context of the multiverse. The problem is addressed by taking the view that the "local" time in our universe should be distinguished from the fundamental time of the multiverse. The multiverse is a closed system, thus it preserves the time symmetry. Our universe is an open subsystem and the event of its birth breaks the time symmetry locally creating an arrow of time in its domain. Since the physical laws in our universe are inherited from the multiverse then it follows that they are time symmetric despite the fact that the whole universe has broken the time symmetry from the moment of the big bang.

An overview of the many arrows of time and their connections is given by Rüdiger Vaas. He discusses different explanations for their origin and focuses on useful conceptual distinctions. Furthermore he suggests a multiverse framework in which the (or our) big bang created the arrows of time. In this framework, the big bang might have originated as some sort of pseudo-beginning in a quantum vacuum that has no direction of time (macrotime) but nevertheless some sort of symmetric microtime. It is even possible that time ends – although paradoxically, it may do so only temporarily. Some recent cosmological models do in fact instantiate such pseudo-beginning and -ending scenarios.

The close correlation between timekeeping devices and physical laws is addressed both in the chapter by Rodolfo Gambini, Rafael A. Porto, and Jorge Pullin and in that by Andreas Albrecht and Alberto Iglesias. Albrecht and Iglesias focus on the general case of the ambiguity associated with the choice of clocks, leading to the immediate implication that we cannot have a fixed set of laws since different clocks would lead to different predictions of the theory. The view taken there is that the problem of clock ambiguity may be bypassed if physical laws emerge statistically from a random time-independent Hamiltonian.

Gambini, Porto, and Pullin present crucial aspects of the impact clocks have on quantum theory, specifically the measurement problem. Devices that measure time, like all quantum systems, are subject to quantum fluctuations. Therefore they are constrained by a fundamental bound on the precision of their timekeeping ability. The intrinsic uncertainty of clocks, given by this bound, makes it unrealistic to expect an accurate, deterministic measurement of time in quantum systems, including physics near the big bang or a black hole.

Martin Bojowald tackles the issue of clock ambiguity and its quantum uncertainty, based on a phenomenological model inspired by quantum gravity, whereby changing clocks is equivalent to a gauge transformation.

Several implications of time's arrow, along with proposals that circumvent this problem, are presented in the chapters by Tom Banks, by Gary W. Gibbons, by Katherine Freese, Matthew G. Brown, and William H. Kinney, and by H. Dieter Zeh. The origin of the cosmic arrow of time is closely related to the origins of the universe by the second law of thermodynamics. According to this law, the universe must have started in an incredibly ordered (low entropy) state in order to be consistent with the observed time's arrow.

The possibility that the origin of the arrow of time is rooted in quantum cosmology is presented in the chapter by Claus Kiefer. This origin would be a consequence of imposing low entropy boundary conditions on the wavefunction of the universe. A low entropy boundary condition is a natural choice since the decoherence process increases the entropy of the wavefunction in an irreversible manner.

Tom Banks advocates applying the holographic principle to the Boltzmann–Penrose question of why our universe started in such a low entropy state, while Freese, Brown, and Kinney provide a concrete model of "phantom bounces and oscillating cosmology" in which the universe is naturally driven through low entropy states at the start of each cycle. Gibbons proposes a sharp form of Thorne's hoop conjecture for the formation of black holes, which relates Birkhoff's invariant to the ADM mass of the outmost apparent horizon.

An interesting question related to time's enigma is the puzzle of why different arrows of time, such as for example, the cosmic arrow determined by the expansion of the universe, the biological arrow determined by (say) human ageing or the thermodynamic arrow determined by the increase in entropy, agree with each other. H. Dieter Zeh makes the case that the "Master Arrow of Time" (the combination of all time's arrows) does not have to be the same as a formal time parameter needed to measure the succession of global states. Zeh also discusses the arrows in both classical and quantum physics, the retardation of various kinds of correlation, the dynamical rôle of quantum indeterminism, and different concepts of timelessness (quantum gravity included).

Reading through the chapters of this book takes us on a fascinating voyage through the diversity of current schools of thought on the very basic question – what is time? A question that remains stubbornly obscure despite centuries of investigation. Time, the entity we are all intuitively wired to acknowledge and take for granted from birth.

As Lord Byron wrote: "Time! The corrector when our judgments err..." Hopefully time will tell which of the judgments presented in this book will stand the test of time.

Time After Time — Big Bang Cosmology and the Arrows of Time

Rüdiger Vaas

Abstract Time, as familiar as it seems to us in everyday life, is one of the greatest puzzles of science and philosophy. In physics and cosmology it is especially mysterious why time appears to be "directed", that is, why there seems to be an essential difference between the past and the future. The most basic known laws of nature do not contain this asymmetry. And yet, several arrows of time can be distinguished – at least ten, in fact. However, it is unclear whether any of them are fundamental or whether others can be reduced to these, and it is not known how the direction of time could be explained convincingly. From the growing but still astonishingly low entropy of the observable universe, it seems plausible that the solution of the mystery is connected with cosmology and an explanation of the big bang. This could require a new fundamental law of nature (which might be related to a particular geometry) or specific boundary conditions (which might be comprehensible within the framework of a multiverse theory). Or it may be that time's direction is fundamental and irreducible, or an illusion and not explicable, but can only be "explained away". It is even more confusing that not all of these alternatives are mutually exclusive. Furthermore, there is a plethora of approaches to explain the big bang. Some models postulate an absolute beginning of time, others an everlasting universe or multiverse in which the big bang is a phase transition, and maybe there are myriads of big bangs. So the low entropy of the observable universe might be a random fluctuation – whereas elsewhere even opposite thermodynamic directions of time may arise. Perhaps the (or our) big bang just created the arrows of time, if it originated as some sort of pseudo-beginning in a quantum vacuum that has no direction of time. Thus it seems useful to conceptually distinguish an undirected microtime and a directed macrotime. It is even possible that time ends – although paradoxically, it may do so only temporarily.

R. Vaas (✉)
bild der wissenschaft, Ernst-Mey-Str. 8, D-70771 Leinfelden-Echt., Germany
e-mail: Ruediger.Vaas@t-online.de

L. Mersini-Houghton and R. Vaas (eds.), *The Arrows of Time*, Fundamental Theories of Physics 172, DOI 10.1007/978-3-642-23259-6_2,

Man is ... related inextricably to all reality, known and unknowable ... plankton, a shimmering phosphorescence on the sea and the spinning planets and an expanding universe, all bound together by the elastic string of time. It is advisable to look from the tide pool to the stars and then back to the tide pool again.

John Steinbeck: The Log from the Sea of Cortez (1951)

Talkin' bout that youthful fountain
Talkin' bout you and me
Talkin' bout eternity
Talkin' bout the big time
Neil Young: Broken Arrow (1996)

1 The Direction of Time

"Time flowing in the middle of the night, / And all things creeping to a day of doom," wrote the British poet Alfred Lord Tennyson. Yet this unceasing stream of time, existing apparently without dependence on its recognition, is perhaps only an illusion – but also a problem. Because the known laws of physics are time-symmetric. So they neither entail nor prefer a direction from past to present.

However, our everyday experience teaches us the opposite. For only processes with a clear direction are observed in the complex systems of nature and culture: blossoms become apples that later decompose; milk drops into black coffee, making it brown; a glass falls from the table and bursts into a thousand pieces. Even cyclic processes of nature such as the seasons or the phases of the moon are parts of irreversible dynamics. Whoever watches mold turning into a red apple, milk drops hopping from a coffee cup, or shards being resurrected into a glass probably would feel like he is in the wrong movie – or simply watching one that is running backwards.

Irreversibility is why – or how – the formation and development of complex structures is much less likely than their decay or something turning into dust and ashes. By use of the concept of entropy this can be quantified physically: it is a measure of a system's degree of disorder. And disorder is much more probable than order. There are, for example, significantly fewer possibilities of molecular combination for a small drop of milk in coffee than for a good mixing. This is why entropy only increases on average, as the second law of thermodynamics states, while the first law expresses the conservation of energy (see [28, 29, 32, 91] for a historical introduction to thermodynamics).

The development of local order does not contradict the second law of thermodynamics. Contrariwise, it creates more disorder within the entire system. In general, entropy does not decrease globally, but can do so locally. Therefore, the formation of complex structures, in other words order, is not impossible, but it occurs only at the expense of a greater amount of disorder in the environment (see, e.g., [65, 66]). Cleaning your desk, for instance, means eating more lettuce, the leaves of which gain their energy from nuclear fusion in the sun – local order increases, yet so does the amount of chaos in the solar system.

So the second law marks a direction of time – or developments in time, which does not necessarily mean the same thing. Yet the second law is not the solution of the problem, but its core. Because all of the known apparently fundamental laws of nature are time-symmetric: they don't include entropy increase; they don't contain a preferred direction of time; they don't differentiate between future and past in principle. This time-reversal invariance means that every macroscopic process could also run in reverse. So why doesn't it in our universe?

This question could be rejected as meaningless if one argues like this: disorder increases with time because we measure time in the direction in which disorder increases. However, this does not solve the problem, because it would still remain unclear why the thermodynamic direction of time exists in the first place. The developments could, after all, also alternate between forward and backward – or not take place at all (see [116]).

It is important to bear in mind that time-reversal invariance and reversibility are not the same but independent from each other and not necessarily correlated (following [3]). Time-reversal invariance is a property of dynamical equations and of the set of their solutions. Reversibility is a property of a single solution of such an equation. Dynamical equations are time-reversal invariant if they are invariant under the application of the time-reversal operator T, which performs the transformation $t \to -t$ and reverses all dynamical variables whose definitions as functions of t are not invariant under this transformation. If $f(t)$ is a solution of such an equation, then $Tf(t)$ is also a solution. These "time-symmetric twins" are temporal mirror images of each other and only conventionally different if no privileged direction of time is presupposed. A solution $f(t)$ is reversible if it does not reach an equilibrium state where the system remains forever. (In classical mechanics, for instance, a solution of a dynamical equation is reversible if it corresponds to a closed curve in phase space.) Time-symmetric laws are, in conclusion, perfectly compatible with asymmetric solutions (see [182]).

2 Ten Arrows of Time

Why do we remember the past, but not the future? For this asymmetry of our experience of time – an irreversibility of many processes and thus the direction of time – Arthur Stanley Eddington [41] coined the metaphorical expression "arrow of time".

It is an open and controversial issue whether there is a single arrow, including and perhaps "guiding" all physical processes, or whether some processes evolve in some sense independently of each other, instantiating their own arrows of time (detailed reviews and elaborations are given, e.g., by [35, 67, 131, 181]).

Conceptually, at least ten different arrows – categories of phenomena that have a direction in time – can be distinguished (see [148]):

- The psychological arrow of time: we remember the past, which seems immutable, but not the future, which isn't fixed for us yet. We experience a "stream" of time that doesn't turn back but moves us from birth to death. The psychological arrow is related to a computational arrow, if cognitive processes are computational – at least partly (omitting issues of phenomenal content aka qualia here).

- The causal arrow of time: effects never precede their causes, and these have coherent structures (at least in classical systems).
- The evolutionary arrow of time: complex natural but also cultural systems are based upon directed developments and often also upon differentiation. Exponential growth can only be observed in self-organizing systems.
- The radioactive arrow of time: exponential growth is confronted with exponential decay of radioactive elements which marks a direction in time as well.
- The radiative arrow of time: electromagnetic radiation diffuses concentrically from a point but never coincides at one point after moving in concentrically from all sides. (This is also true for sound waves, or for waves that result from a stone being thrown into water, or for the assumed gravitational waves emitted by rotating, collapsing, or colliding massive bodies.)
- The thermodynamic arrow of time: the entropy of a closed system maximises, so the system seems to strive for its thermodynamic equilibrium. For example, coffee cools down to ambient temperature and milk drops that have been poured into it don't stay together but disperse evenly.
- The particle physics arrow of time: the decays of certain particles, the neutral K mesons (kaons) and B mesons, and there antiparticles lead implicitly to the conclusion that there is an asymmetry of time because these decays break other symmetries. (More precisely, some processes governed by the weak interaction violate time reversal T, but can also be subsumed under time-reversal invariance nevertheless, because T-violation is compensated by an application of a unitary CP-transformation, and according to the CPT-theorem the combination of charge conjugation C, parity transformation P, and time reversal T is conserved.)
- The quantum arrow of time: measurements – or interactions with the environment (quantum decoherence) in general – interfere with a quantum system which realizes all possible states in superposition, and lead to only one classical state being observed. This so-called collapse of the wave function (if it really happens) describes, for example, why Erwin Schrödinger's infamous cat is not (observed as) dead and alive at the same time. Instead of collapsing, reality could also "split" into different parallel universes that would henceforth be independent of each other, so that all alternatives were simultaneously realized – in one world the cat is dead and in another one it is alive.
- The gravitational arrow of time: gravity forms structures, for example galaxies and stars, from tiny density fluctuations within the almost homogeneously distributed primordial plasma of the early universe (Fig. 1). Gravitational collapse can even create black holes. They are "one-way streets" of matter, places of highest entropy, and perhaps even irreversible annihilators of physical information. This arrow is also called (or subsumed under) the fluctuation arrow [70].
- The cosmological arrow of time: space has been expanding since the big bang.

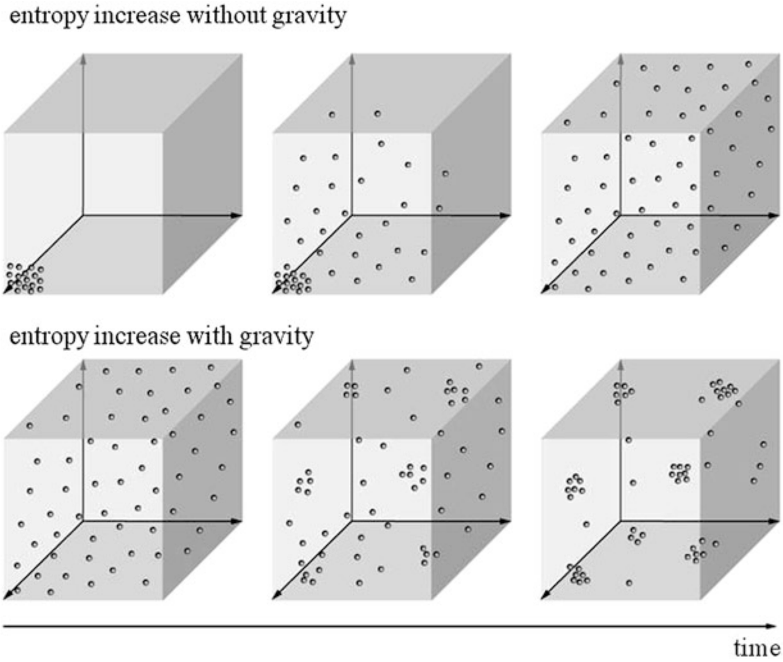

Fig. 1 Growing disorder. Entropy – the physical measure of a system's disorder – can only increase statistically in the course of time. For this reason it even defines in a way the direction ("arrow") of time. If a gas bottle is opened up in empty space, the gas molecules soon spread evenly throughout the entire volume – then, thermodynamic equilibrium is reached as a state of maximum entropy (top). Yet in a large space such as the early universe, gravity creates local concentrations of the originally almost homogeneously distributed gas (bottom) – and this was how stars and galaxies formed. An increase of entropy follows from this gravitational effect, which was not taken into consideration for a long time. There is still a debate about whether there can be a thermodynamic equilibrium, a "heat death", in an expanding space, and how the total entropy in the universe can be usefully defined at all.

These ten temporally directed processes seem more or less unrelated to each other at first glance. Yet given that at least today all arrows point in the same direction, it seems natural to search for a primordial, super, or master arrow of time that all the others could be ascribed to. The particle physics arrow of time, the cosmological arrow of time, and the thermodynamic arrow of time are likely candidates.

The thermodynamic arrow might be responsible for the psychological and the evolutionary arrow of time (cf. [68]; but see [94] for an argument against the correlation of thermodynamic and psychological or computational arrows, respectively). Entropy can also be defined for black holes and hence for gravitational processes (see [85, 113]). Causality is a difficult issue (see, e.g., [128, 129]) taken by some to be subjective or as a logical relation, ultimately, and thus reducible to other arrows (see [117]), but as the "cement of the universe" by others (see, e.g., [92]); thus perhaps causality is not only a pragmatic consideration but grounded in processes governed by conservation laws such as conservation of mass-energy, linear and angular momentum during transmissions of energy and momentum.

Still mysterious is the origin and implication of the particle physics arrow [15]. Joan A. Vaccaro [172] argued that processes which violate T-symmetry induce destructive interference between different paths that the universe can take through time. The interference eliminates most of the possible paths except for two that represent continuously forwards and continuously backwards progress in time. Data from accelerator experiments allow the distinction between the two time directions and indicate which path the universe is effectively following. Thus T-violation might have large-scale physical effects that underlie the unidirectionality of time.

There have also been controversial discussions about whether the arrow(s) of time in an evolving closed universe will reverse in the collapsing phase [35, 54, 55, 67, 71, 72, 82, 107, 112, 113, 181]. Perhaps the different arrows of time reduce to the cosmological arrow, so in some sense the direction of time would switch at the maximum size of the finite universe, when the expansion turns into contraction. Sometimes it has been argued that even the psychological and thermodynamic arrow of time would run backwards (from the perspective of the expanding stage), and observers would still believe they were living in an expanding phase. In a quantum cosmological framework, however, everything with classical properties is destroyed in the maximum stage, due to quantum interference, and the big bang and big crunch are ultimately the same, amusingly called the big brunch [82, 181].

3 Four Kinds of Answers

Where does the asymmetry of time – or at least the processes in time – originate if most laws of nature are time-reversal invariant and thus do not prefer a direction in time? Basically, four kinds of answers can be distinguished [148]:

- Irreducibility. The direction of time is not a derivable phenomenon but an essential attribute of time: then time simply passes and is independent, for example, of entropy. Many philosophers share this opinion. Tim Maudlin [98], for instance, defends it and accuses skeptics of only being able to argue for time symmetry because they already presuppose it. However, this objection might be reversed and Maudlin could be accused of not admitting the problem in the first place.
- Laws. Perhaps there is a fundamental, but still unknown law of nature that is time asymmetric. Accordingly, Roger Penrose [112] hopes that such an arrow of time follows from a theory of quantum gravity that unites quantum theory and the general theory of relativity. This might also explain the mysterious collapse of the wave function that many physicists assume. Therefore quantum theory would have to be modified in such a way that it contains a time asymmetry. Then the past could be calculated from a future perspective but not the other way round. This possibility would help historians to gain an advantage over physicists. Other

researchers, Ilya Prigogine [118] for instance, localize arrows of time in the peculiarities of complex systems far from thermodynamic equilibrium, which are postulated to have special laws.

- Boundary conditions. Most physicists assume that the irreversibility of nature is not based upon time-asymmetric laws but is a result of specific, perhaps very improbable initial or boundary conditions (cf. [4, 131, 181]). The problem would thus be shifted to the origin of the universe and accordingly to models of quantum cosmology, though there is no consensus about the nature and form of these boundary conditions. A subset of such explanations are proposals of cyclicity. Here the universe oscillates through a series of expansions and contractions (e.g. [6, 27, 50, 140]) and/or evolves through a perhaps infinite series of big bangs (e.g. [113], see below). Real cyclic models, which do not shift the problem of time's arrow into the infinite past, have to show how the entropy created in each cycle is destroyed or diluted before or within the subsequent big bang, in order to reset the stage for the next oscillation. Therefore a decrease of entropy or entropy density must be explained.
- Illusion. If time is not objective – a property of the world or at least of some of its objects or their relations – but subjective, physicists are searching for an explanation in the wrong place. Immanuel Kant assumed time to be a pure form of intuition or perception, inherent in the human mind, a kind of transcendental requirement or pre-structure for the possibility of experience itself, hence nothing that belongs to the things in themselves. He claims that "time and space are only sensible forms of our intuition, but not determinations given for themselves or conditions of objects as things in themselves. To this idealism is opposed transcendental realism, which regards space and time as something given in themselves independent of our sensibility" ([80], A 369). Other philosophers suspect time of being a construct of consciousness or of the grammar of our language. There are also powerful, but controversial arguments from physics, especially relativity and quantum gravity, emphazising that there is no time independent from space, only a spacetime unity, or that a time parameter does not even appear in the fundamental equations of quantum gravity (such as loop quantum gravity) or quantum cosmology (especially the Wheeler–DeWitt equation) (see, e.g., [12, 84, 125, 181, 182]). So perhaps the arrows of time do not even exist in the world as such. If the entire history of the universe is there as a whole or unity, time would be a mere illusion in a certain sense.

Some of these accounts are mutually exclusive, others are not. For example time could be an illusion (or emergent), but an asymmetrical block universe (if, say, the big bang has much lower entropy than the big crunch) would still deserve an explanation, which might consist in a specific boundary condition or law. Or if time is fundamental (as, e.g., [31, 102] argue), this might be represented by a law too. Or perhaps specific boundary conditions are really an instantiation of a special law, as suggested by Stephen Hawking [74].

4 Fundamental Issues

It is a deep conceptual, physical, and even metaphysical question whether time is fundamental or not. What does this mean, and how can we know about it?

It is not clear from a conceptual point of view whether the direction of time is a necessary feature of time. If so, and if time is fundamental, then the arrow of time is fundamental too. In this case the chances of finding a deeper understanding, or at least a testable explanation in physics or cosmology, are slim. Time as a fundamental entity, as well as its arrow(s), could of course be represented as a fundamental parameter in a future fundamental theory, but this would be no derivation or explanation, just an assumption. Ultimately, time – and the arrow of time – would remain a mystery. This might very well be the case. However, methodologically, it can and should not be a premise or limit of research. On the contrary, scientists and philosophers alike should try to reduce and/or explain time – and proceed as far as they can get. Even wrong explanations are better than no explanation at all, because they can be revised and improved. And their errors may teach useful lessons nevertheless. If an explanation doesn't work, it could still tell us something new, if it is possible to understand why it doesn't work.

Note also that time could be fundamental whether or not it has a beginning. If time originated with the big bang, there was no "before". On the other hand time might be eternal, thus preceding the big bang, which is compatible with a multiverse scenario producing countless big bangs and, hence, universes. But the view that time is emergent is also consistent with either an absolute beginning or temporal eternity within or without a multiverse.

If time is fundamental, this doesn't imply logically that the arrow of time is also fundamental. Perhaps time's direction requires additional assumptions, such as causality or specific initial conditions, which might not be fundamental and could be explained (or they are purely accidental and therefore not further explainable).

For example, the fundamental theory might include a basic time parameter but still not tell us why the entropy of the universe is as low as it is, nor why time's direction could not change. The fundamental theory might be time-symmetric nevertheless – just as classical mechanics, the theory of electromagnetism, general relativity, quantum mechanics, and quantum field theory are. Alternatively, there could be a fundamental, even eternal time within a multiverse scenario where different universes or parts of the multiverse have different directions of time. Or microtime may be fundamental while macrotime (including an arrow, see [154,162], and below) is not; thus there may be places without local or even global arrows of time – as there could exist islands of reverse arrows (cf. [131–134], but against this [181]). Perhaps the far future empty universe will approximate to such a timeless place (cf. [170]), or there may already be localized regions somewhere within the universe, or there was such a state before the big bang, e.g., a quantum vacuum. Of course such conceptual possibilities are not solutions of the problem, just surveys, and conclusions must be supported by scientific arguments.

If there is no (or no non-reducible) time parameter in the fundamental theory, which is not known yet, one might argue that time is not fundamental – following the view that metaphysics should be determined or framed by our best scientific theory. This reasoning is controversial. But if one accepts the nature of time as a (at least partly) metaphysical issue at all, then attempts to understand it should be in accord with the best scientific theories. And if the fundamental theory contains no time, as some approaches in quantum gravity already suggest, time might be "emergent" or illusionary indeed. If so, ordinary time – or its many aspects – can (or must) be explained in a certain sense. And then there are good chances that the arrow of time can be explained too – at least approximately.

This explanation need not necessarily be a physical or cosmological one, by the way. Perhaps "time" has something to do with how we are practicing science, that is predict and retrodict events and facts. But this is based upon our everyday thinking, how we deal with our experiences and how we order sensations and intentions; it could have been simply advantageous in the evolution of our cognition and behavior (perhaps even a kind of useful illusion such as believing in free will or deities, see [149, 160]). Thus it might turn out that it is sufficient to take time as an ordinary-life concept, a way to describe and handle sensations and actions, to characterize it phenomenologically, and, perhaps, to search for a neuropsychological (or even neurophysiological) explanation. If so, the riddles of time would not be a genuine part of physics, only inherited by physics or transformed into it, but ultimately solved by cognitive neuroscience (see, e.g., [119, 120, 147]). In this respect, time might even be fundamental to us, together with space, that is "pure forms of sensible intuition, serving as principles of a priori knowledge", as Kant ([80], B 36) put it, and hence experimentally opaque, but for practical reasons transferred as a parameter into scientific theories. Thus time could be both fundamental (for observers) and an illusion (not existing mind-independently) – and even emergent (e.g., arising in complex neural networks of cognitive systems). To avoid conceptual confusion it is therefore important to clarify notions such as "fundamental", "emergent", "reducible", "illusionary", etc. in respect of the scope of application.

Though the concepts of time and its direction are indisputably important for our cognitive setting, it would require strong arguments assuming it is sufficient to reduce questions regarding the arrows of time in physics and cosmology simply to cognitive neuroscience or even philosophical phenomenology. It is trivial that science requires scientists, but it would be a non sequitur to claim because of this that there are no features independently from scientists or conscious states and events in general. It would be very surprising if scientific explanations end in or lead to scientific minds, rather than starting from them.

5 Gravity, Entropy, and Improbability

The second law of thermodynamics results – at least phenomenologically – from there always being more disordered states than ordered states. This can be illustrated by a box with many pieces of a puzzle. There is one and only one arrangement in which the pieces create a picture. Yet there is a high number of combinations in which the pieces are disordered and do not form a picture. This is similar to the molecules of stirred milk in a cup of coffee: theoretically they could agglutinate into a drop; in practice they never do because this is so improbable. The reason for such extremely low likelihoods is not represented by laws of nature, however, but by the boundary conditions, respectively the initial conditions. And it is these that pose a conundrum.

So one can argue like this (see [177]): Why does the thermodynamic arrow of time exist? Because the present entropy is so low! And why is it so low? Because it was even lower at earlier times!

This explanation is, however, as elegant as it is insufficient. Because it only shifts the problem, relocating it to the remote beginning of our universe. Yet the big bang 13.7 billion years ago lies in a dark past – and this is not just meant metaphorically. There was no light until 380,000 years after the big bang, when the universe had cooled down enough to release the cosmic microwave background radiation that we can measure today. Given that this radiation is extremely homogeneous – aside from tiny fluctuations in temperature on the order of a hundred thousandth of a degree – matter must have been extraordinarily uniformly distributed at this early epoch and in thermal equilibrium with the radiation. (Dark matter, if it exists, does not interact electromagnetically, and would have been 10 to 100 times more concentrated.)

The spectrum of the cosmic background radiation today almost perfectly resembles the electromagnetic radiation of an idealized black-body in thermal equilibrium with a temperature of 2.725 K (with an emission peak at 160.2 GHz). This might appear paradoxical at first, given that such an equilibrium is often assumed to be the maximum of entropy – like the heat death of the universe that physicists in the 19th century imagined to be the bleak end of the world, consisting ultimately only of heat and perhaps homogeneously distributed particles, if there are any that cannot decay.

Yet appearances are deceptive: the homogeneous fireball of the early universe did not have a high, but a very low entropy! Because in the balance gravity must not be ignored – something that was not recognized for a long time. And gravity is working in the opposite direction: clumping, not homogenizing. So at large scales homogeneity doesn't show a high entropy, but contrariwise a very low one, because gravity's part of the entire entropy here is very low. The strongest "concentrations" of gravity, black holes, are also the biggest accumulations of entropy. Physically speaking, gravitational collapse leads to the greatest possible amount of disorder. The entropy of a single black hole with the mass of a million suns (such as the one at the galactic centre, for example) is a 100 times higher than the entropy of all ordinary particles in the entire observable universe. Yet the homogeneous cosmic background radiation and further astronomical observations very clearly show that black holes did not dominate the very early universe, and this has remained so until today.

This extreme uniformity of matter distribution and the "flatness" of our universe's spacetime geometry themselves appear almost as a miracle. Penrose [111, 112] was the first to recognize and even quantify this. Compared to all possible configurations of matter and energy in our universe, the actual state is extremely improbable.

Penrose estimated it to be a mere $1{:}10^{10^{123}}$, more recent data imply circa $1{:}10^{10^{122}}$ [85]. This double exponent is unimaginably huge. It has so many zeros that it would, if printed in the format of this book, amount to a stack that were considerably higher than the diameter of our observable universe. Thus a universe filled with black holes is much more likely than ours. Yet we don't observe such a black hole entropy dominated universe – and we couldn't even live in one. Viewed in this light, $1{:}10^{10^{122}}$ becomes a requirement for our existence.

One might argue therefore on the basis of the weak anthropic principle [13, 153] that we should not wonder about the low entropy, because if it were much higher, we could not exist and there would be no one to wonder about it. So low overall entropy is certainly a precondition for complex life. However, a much higher overall entropy would suffice, making such an argument very unconvincing. Therefore the anthropic principle is insufficient for a comprehension of time's direction, because the observable universe is much more ordered than would have been necessary for human existence. To be more accurate: the probability of our entire solar system including earth and all its life-forms popping out of coincidentally fittingly arranged particles might only be $1{:}10^{10^{85}}$ – but this is overwhelmingly more probable than the $1{:}10^{10^{122}}$ for the entire observable universe. So the anthropic principle is not helpful here: neither as a mere tautology stating a necessary condition for life nor as a selection criterion for a universe that makes life possible within a multiversal realm of possibilities, because even if there were $1{:}10^{10^{122}}$ universes differing in their initial conditions, this would not render the actual value of entropy in our universe plausible.

6 Beyond the Big Bang

We exist in a world full of order that is friendly to life in the thermodynamical sense because the big bang was supremely "orderly". And, as most scientists are convinced by now, this is exactly the reason why the universe runs like a clock – indicating a clear direction of time. But what was it that wound up the cosmic clockwork? How did this supremely special big bang come about? What caused the low entropy of the early universe?

Some 13.7 billion years ago the observable universe evolved from an extremely hot and dense region smaller than an atom which expanded enormously. While the aftermath of this big bang is both theoretically and empirically well established, and to a large extent understood, it is still a mystery as to how and why the big bang occurred at all. Was it the beginning of space and time, or only of matter? If it was a transition, what came before? If not, how could "everything" appear out of "nothing"? And was it a singular event or one of perhaps infinitely many. Do other universes also exist, and did they or will they interact with our own? These are difficult questions and controversial issues – but no longer beyond the scope of science. In modern quantum cosmology a lot of competing scenarios are being pursued [167]. They open up the exciting prospect of going “beyond” the big bang and even of finding a physical explanation for it.

“Ad fontes” (“to the sources”) – this humanist slogan from the early modern period could be fitting for today’s physicists too: in order to understand the direction of time, they also have to discover the origin of time. Yet the early days of the universe appear as incomplete, misleading, and dark, as historical sources often are. Considering the far longer periods of time, it is surprising that anything at all should still be preserved – and that cosmologists can partly decipher it. Indeed, the observable universe might have “forgotten” much of the information it held in its primordial times. This could be a result of cosmic inflation (insofar as it actually happened). The result of this huge expansion of space is that hardly anything remains in the observable universe from the time of inflation – if inflation had a beginning at all (and has not been going on since all eternity), something that most cosmologists presume indeed. But even in this case, our universe might have separated from the inflationary epoch at a randomly late point. Less than a hundred volume doublings would have been sufficient to cover all tracks from the time before inflation. Cosmic inflation has even been assumed to be the source of the low entropy of our universe [4]. Yet it seems that inflation alone could not have accomplished all of this (e.g., [100]). On the other hand, it might at least be the key to the door of such a deeper explanation that would have to make the initial conditions of inflation understandable – something that can be criticised as yet another shift of the problem however.

In the end, the breakthrough to a deeper understanding will be up to the theoreticians – in the form of a theory of quantum gravity that would have to be confirmed howsoever. The challenges are enormous, and the consequences are as yet unclear. Even our old companion time will probably not be left unblemished. It seems to dissolve entirely in the noise of the smallest scales of nature where there are no longer any clear, regular oscillations, and hence also no “clocks” (see [84]). The disturbing consequences of the theory of relativity – which reduced time to a “fourth dimension” and merged it with space into a unity [115] – cannot be reversed in a quantum theory of gravity, but are here to stay. General relativity implies that there is no background spacetime – no stage where things move autonomically, without affecting spacetime. Hence, there is no “time” that everything could flow along. This seems to be even more true for a theory of quantum gravity. Here the notion of a spacetime continuum breaks down at the Planck scale, turning lengths and time intervals into quasi-discrete entities. Perhaps the world must be described without a concept of time on its fundamental level [83, 123].

Nevertheless the big bang still appears special, and the arrows of time, whether fundamental or not, deserve an explanation. This might even reach beyond the big bang. Actually the big bang was not necessarily the absolute beginning of everything. Whether it was or whether it happened, on the contrary, as a phase transition – for example a “bounce” of an earlier, contracting universe or an accidental fluctuation within a quantum vacuum – is an open and very controversial issue (see below). But in principle the fluctuation or bounce scenario is a promising candidate for a dynamic origin and, thus, explanation, of the arrow of time.

Furthermore, if the big bang was not the beginning of everything but a phase transition, one need not ask how something came out of nothing (which is of course a different question than why there is something rather than nothing): the big bang then was not something that sprang into existence ex nihilo, and nor did spacetime or energy or the laws of nature. It is also meaningless to ask why the entropy was so small at the beginning, if there was no ultimate beginning at all. Nevertheless this question reappears in a modified form: Why was the entropy so small at the bounce or at the beginning of the fluctuation? If it had been large, the big bang would not have produced the smooth, low-entropy universe which is still observed today, but a chaotic mess.

7 Big Fluctuation

If the big bang was a fluctuation, the special low entropy of the universe could have originated as a pure accident (and therefore could be explained away).

But even if our observable universe were only a coincidentally developed island of order in a much greater ocean of chaos – a statistical fluctuation, as Ludwig Boltzmann deliberated as early as 1895 – then it would still be incomprehensible why this fluctuation is so persistent (Fig. 2). After all, about 13.7 billion years have passed since the big bang. But it appears to be much more probable for the spontaneous fluctuation to have arisen only last Thursday or a few seconds before this very moment right now – with all the pseudo-traces of an alleged past: the memories of earlier tax declarations and children's birthdays, the fossils of dinosaurs, the meteorites from the beginning of the solar system and the cosmic background radiation from the aftermath of the big bang itself. In a nutshell: such a bogus-universe – or only a single brain in which such a pseudo-world manifests itself – should arise overwhelmingly more frequently simply by chance than a highly structured, ordered space of at least 100 billion light years in diameter. This often disregarded objection was already made (roughly) by Carl Friedrich von Weizsäcker [176] in 1939 and reappeared in modern dark energy cosmology as the problem of the Boltzmann brains [39, 90, 161]. To give some thermodynamical numbers of entropy fluctuations in a de Sitter background, the probability of our observable universe, $1{:}10^{10^{122}}$, is extremely tiny in contrast to a spontaneous ex nihilo origination of a freak observer, perhaps $1{:}10^{10^{21}}$ for the smallest possible conscious computer, and between $1{:}10^{10^{51}}$ and $1{:}10^{10^{70}}$ for a "Boltzmann brain" [36]. (Thus, there is a controversial discussion going on about wrong assumptions underlying those kinds of estimates – not because many scientists believe that such a solipsistic illusion is true, but because these probabilities indicate possible errors in cosmological reasoning and deep difficulties of multiverse models, especially the measure problem in inflationary cosmology.)

Of course, a virtue can be made out of necessity. Sean Carroll and Jennifer Chen [30] did just that. They argue that our universe really is a mere fluctuation

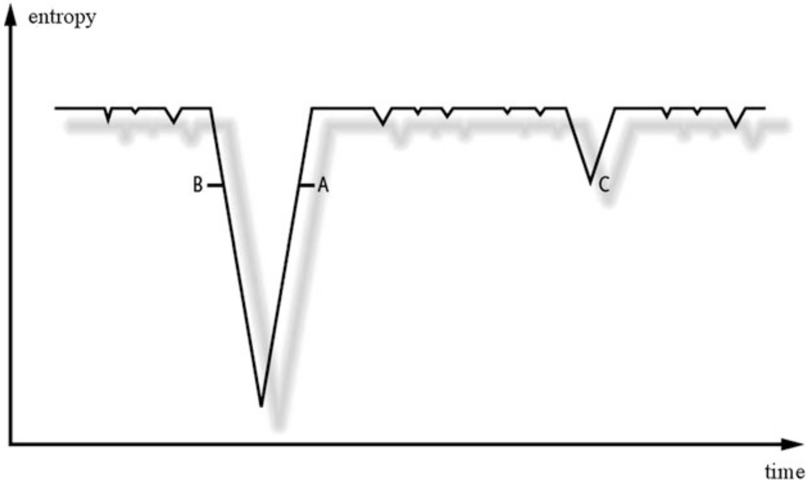

Fig. 2 Order from chaos: When a system is in a state of maximum disorder – i.e., entropy – then temporary "islands of order" and thus local directions of time (point C) develop by means of chance processes over long periods of time. That's why there have been recurring speculations about the entire observable universe being such an island in the midst of chaos. Intelligent observers could only live within one of these "entropy gradients" (point A). Yet there are two fundamental difficulties with such a viewpoint. First, it would be much more likely for everything around A to have originated out of chaos only very recently (as in the case of C) – but then most of what seems to have happened in the past would be a mere illusion. Second, life-forms in the vicinity of B would experience the direction of time exactly in reverse to A.

among myriads. This is possible if the entirety of empty space, taken as a quantum vacuum, contains even more entropy than isolated black holes that only have the maximum entropy within a specific volume. Such an (eternal) accelerating expansion of space, driven by the still mysterious dark energy, could indeed entail an even higher entropy than black holes. Yet such a vacuum must produce random quantum fluctuations again and again. Some of them become huge because of inflation, until they deflate entirely due to the perpetual expansion caused by dark energy. And such cycles, according to Carroll and Chen, are more likely than random fluctuations of dinosaurs and bogus-universes. Thus, in an infinite future, time might not be a problem. Eventually, anything could spontaneously pop into existence due to quantum fluctuations if spacetime is eternal. They would mostly result in meaningless garbage, but a vanishingly small proportion would contain people, planets, and parades of galaxies. This book will also reappear again (a modern version of the well-known philosophy of eternal recurrence, which has many other versions in current cosmology too, see [168]). And this kind of quantum resurrection might even spark a new big bang. According to Carroll and Chen, one must be patient, however, and wait some $10^{10^{56}}$ years for another recurrence of our observable universe (if a de Sitter vacuum with a positive cosmological constant Λ is the "natural ground state"). Our whole universe might be such an island in a Λ-sea,