

Weakening Gravity's Grip on the Arrow of Time

Maulik Parikh^{†*}

[†]Inter-University Centre for Astronomy and Astrophysics (IUCAA)
Post Bag 4, Pune 411 007, India

Abstract

The future differs from the past: it has more entropy. No theoretical framework including inflation has yet provided a dynamical origin for this elementary fact, the thermodynamic arrow of time. I argue that by weakening the strength or range of gravity at early times, one can find a natural way to obtain the smooth conditions present in the early universe.

*email: parikh@iucaa.ernet.in

Time's Arrows

The nature of time has perplexed philosophers and physicists throughout the ages, triggering furious debate and generating papers, conferences, grants — even essay contests. Does time exist? Is it emergent? Does it flow? What makes this particular moment the present? What (or when) was time before the Big Bang? One has to be careful. Not all questions that are syntactically correct in English are meaningful, let alone profound, inquiries into physics; partly because we lack a deep understanding of the structure of spacetime, conjectures about the nature of time walk the thin line between the abstruse and the absurd. Here I will turn to what ought to be a safer query: why does time seem to have a direction?

Perhaps no aspect of time is more self-evident than its directionality. Indeed, every macroscopic closed (or almost closed) system looks qualitatively different in the future than in the past. Gases initially filling half a box go on to fill all of it, people grow older, and stars eventually run out of fuel and die. The reverse never happens. One could reasonably have hoped that such a simple observation would be a trivial consequence of the laws that govern the behavior of gases, people, and stars: the effective field theory comprised of the standard model of particle physics and general relativity. But contrary to this expectation, these great theories treat the future and the past with almost perfect indifference; the origin of the *arrow of time* remains obscure.

What exactly do we mean by an arrow of time? Last year, an amusing puzzle appeared in the online version of the *New York Times*. Filmmaker Errol Morris wanted to know which of two nearly identical photographs, taken a few hours apart in the aftermath of a battle in the Crimean War, was taken first. Online readers posted hundreds of ideas, based on lighting and shadows or assumptions about cannonball kinetics or even guesses about the motives of the photographer, but the convincing — and correct — solution relied on the second law of thermodynamics: the picture in which a number of small pebbles had higher entropy was taken later. This anecdote illustrates a general principle: given two instantaneous snapshots of a system, an arrow of time is *defined* by an ordering of the snapshots.

Using the notion of sequential snapshots, one can tease out several different kinds — indeed a whole quiver's worth — of temporal arrows. First and foremost, there is the thermodynamic arrow of time, which points in the direction of increased disorder. Then there is the electromagnetic arrow of time, to do with the fact that we choose retarded rather than advanced potentials when solving Maxwell's equations. There is the psychological arrow of time, pointing from the past we remember to the future we don't. And then there is the cosmological arrow of time, in which snapshots are arranged according to the size of the universe. (There are others too.) These arrows could, a priori, have been independent of each other. We would then have had to explain not only how these arrows arise from the underlying dynamics, but also the secondary question of why the arrows have their mutual relation to each other. In fact, however, most of the arrows of time are not independent. For example, recording information increases overall entropy and so the ordering in which entropy increases is the same as the one in which memories get fuller; hence, the psychological and thermodynamic arrows of time are aligned. It seems likely that all the other arrows of time (with the possible exception of the cosmological arrow) are fixed by the thermodynamic arrow of time. But the thermodynamic arrow itself has no easy explanation. In this essay, I will describe a novel idea for obtaining the thermodynamic arrow dynamically, in a “natural” way, if one is allowed to make a few modifications to gravity in the early universe [1] .

Entropy

Things fall apart; the centre cannot hold;
Mere anarchy is loosed upon the world
— W. B. Yeats

It was Ludwig Boltzmann who realized that the extraordinary diversity of time-asymmetric phenomena could be encapsulated by a single potent equation:

$$\frac{dS}{dt} \geq 0 . \quad (1)$$

In words: entropy increases with time. Here entropy can be defined either classically, as the logarithm of the volume (Γ) of phase space, or quantum-mechanically, from the reduced density matrix (ρ) or as the logarithm of the number (\mathcal{W}) of microstates in a given macrostate:

$$S = k_B \ln \Gamma = -k_B \text{Tr} \rho \ln \rho = k_B \ln \mathcal{W} , \quad (2)$$

the last of which is the epitaph etched on Boltzmann's gravestone. Boltzmann's crisp statement of entropy increase supplies the thermodynamic arrow of time in the sense of the previous section: given two snapshots of a system (more precisely, of its phase space) at different instances, we say that the one with the greater entropy comes later. Indeed, it provides a rule for when a thermodynamic arrow of time exists at all:

$$S_{\text{actual}} < S_{\text{maximum possible}} . \quad (3)$$

Whenever the actual entropy is strictly less than the maximum entropy, the system will evolve towards maximal entropy. Things will change in time, becoming more disordered; an arrow of time is induced. Conversely, when the system has already reached maximal entropy, nothing much happens and there is no arrow at all. This makes sense intuitively; a film clip of a gas that has already filled up its volume will have nothing further happening and will look the same played forward or back. A maximal entropy system has no arrow of time.

The second law of thermodynamics, rooted as it is in statistical mechanics rather than in any particular kind of dynamics, has sweeping, awesome power. Rather than separate laws explaining why glass shatters but never unshatters, or why eggs break but never unbreak, or why black holes decay but radiation doesn't collapse, we have one all-encompassing law. One merely has to compare entropies to find the correct sequence of events. The low entropy state (wine glass, intact egg, black hole) comes before the high entropy state (broken shards, egg mess, Hawking radiation).

However, the second law of thermodynamics is only true on average. Small deviations from equilibrium can and do happen, and Boltzmann was able to quantify the likelihoods of such fluctuations and the time-scales over which they could be expected. In essence, the probability of such a fluctuation decreases exponentially with its departure from maximal entropy:

$$P \sim \exp(\Delta S) . \quad (4)$$

The utility of such considerations can be illustrated by one example. Take protein folding. It had been thought by some biologists that proteins fold into their chosen conformation

(shape) purely by chance. Yet entropy counting tells us that this cannot be true: the time taken for a protein to find its correct conformation would exceed the lifetime of the universe, an observation sometimes known as the Levinthal paradox. We therefore conclude that the process by which proteins fold must be directed, not random.

But Boltzmann went further. He tried to find a microscopic underpinning for the rise in entropy. With his H -theorem, he tried to show that under what he regarded as generic properties (the molecular chaos assumption), a system would invariably tend to greater entropy. And yet, how could that possibly be? If the underlying dynamics is symmetric under the time-reversal operator, \hat{T} , then it must equally be true that entropy increases towards the past! And so it turns out. Indeed, the statement that \hat{T} is a symmetry can be formally written as

$$[\hat{H}, \hat{T}] = 0, \quad (5)$$

where \hat{H} is the Hamiltonian. The Hamiltonian is of course the time evolution operator and hence, when (5) holds, any property derived from evolving the system forward in time must also be obtained when the system is evolved backwards in time. In particular, if the entropy increases to the future, it must perforce increase to the past.

In the end, the second law emerges not from the dynamics but from a choice of very special initial states, namely those with low entropy in the past. The vast majority of these have higher entropy in the future, which is why we always see entropy increase. But, manifestly, the breaking of time reversal symmetry arises from a theoretically unsatisfying choice of special initial conditions. Such low entropy initial conditions are “unnatural” in the sense that with any straightforward measure they occupy an exponentially small (and unstable) region of phase space. And if regarded as fluctuations from equilibrium, they are very unlikely. Of course there is no reason at all to suppose that the universe was ever in equilibrium. So it isn’t quite right to say that low entropy initial conditions are *unlikely*, hence the euphemism “unnatural.” Nevertheless we would like an explanation for how low entropy initial conditions might have come about. The situation is akin to deducing the number of generations of quarks and leptons; there is no particular reason for saying that three generations is unlikely, but it is still a mysterious fact that we would like to be able to derive.

The Usual Suspects

In seeking an explanation for the low entropy initial conditions of the universe, several false starts have been made. One easily refutable proposal is the anthropic argument for the arrow of time. The argument goes like this. A system with maximal entropy will occasionally undergo spontaneous fluctuations to lower entropy configurations; it is not impossible, if one waits long enough, to find all the gas molecules in a room congregating in one corner. The same argument could be applied to the universe as a whole. Imagine an infinitely long-lived universe in an equilibrium state of maximal entropy. A spontaneous fluctuation could put the universe in a lower entropy state, at which point, since S would be less than S_{\max} , an arrow of time would automatically arise. All we have to do is wait. However, it is well known that this naive argument suffers from a fatal flaw. As (4) indicates, the probability of a decrease in entropy falls off exponentially with the decrease; small reductions in entropy are thus far, far more likely than large ones. It would be much more probable that we would decrease to just

the amount of order necessary to give rise to us, than to one that also produced the universe around us. For example, a fluctuation from equilibrium with initial density perturbations $\delta\rho/\rho \sim 10^{-4}$ is vastly more probable than the observed $\delta\rho/\rho \sim 10^{-5}$, and would not rule out observers like us. Taken to its logical extreme, it would be even more likely that we are a Boltzmann brain, the minimal sentient being capable of imagining itself living in an ordered universe.

A more serious candidate for the origin of time’s arrow is the theory of inflation. Inflation involves an accelerated expansion of space, which stretches out any inhomogeneities before they can collapse. Thus inflation quickly leads to a configuration that is smooth and homogeneous. This is actually a low entropy configuration once gravity is taken into account. Despite this, inflation does not ultimately help with the arrow of time problem, the reason being that the initiation of inflation itself requires extremely special initial conditions. In the simple case of a single scalar field ϕ in a potential $V(\phi)$, the equation of state parameter is given by:

$$w = \frac{\frac{1}{2}(\dot{\phi}^2 + (\nabla\phi)^2) - V}{\frac{1}{2}(\dot{\phi}^2 + (\nabla\phi)^2) + V}. \quad (6)$$

For acceleration, we need $w < -1/3$, which requires the potential to dominate over “kinetic” energy. Since spatial derivatives contribute to the latter, inflationary expansion requires a patch of spacetime to contain an extremely smooth scalar field — a condition that is in fact more special than the subsequent configuration the dynamics is meant to explain [2]. Another way to see that inflation does not help is by considering a generic state of the sort likely to have come about shortly after reheating. Most such states do not “un-reheat” when evolved back in time. Die-hard proponents of inflation as an explanation for time’s arrow counter that in chaotic inflation scenarios one is bound to find some region somewhere in space where the field has the required special homogeneous profile. But this argument is just the anthropic argument in disguise. In particular, the amount of homogeneity required to give just a few e-folds of inflation is much less than the amount required to produce the 60+ e-folds that must have happened to give rise to our own universe.

Gravity

Finally, we turn to gravity. Gravity affects the discussion of entropy in at least four ways. First, as a long-range attractive force, gravity leads to Jeans’ instabilities: gravitational collapse. Thus one has to be careful about using results from equilibrium thermodynamics [3]. Second and relatedly, a system with gravity evolves not to a homogeneous configuration but to a clumpy one; see Fig. 1. This point will be crucial to our proposed solution. Third, the presence of black hole horizons adds an enormous contribution to the total entropy. Each black hole has entropy equal to a quarter of the area of its horizon, measured in Planck units.

$$S_{\text{matter+gravity}} = S_{\text{matter}} + \sum_i \frac{A_i}{4G_N\hbar} \quad (7)$$

And fourth, the presence of a positive cosmological constant implies the existence of a de Sitter horizon which too, has its entropy, also proportional to the area of the horizon. The

area of a pure de Sitter horizon is greater than the area of de Sitter space with something in it, thus favoring the emptying out of de Sitter space. The existence of a true positive cosmological constant makes the arrow of time problem even harder than it already is; here I will simply assume that dark energy is not a real cosmological constant.

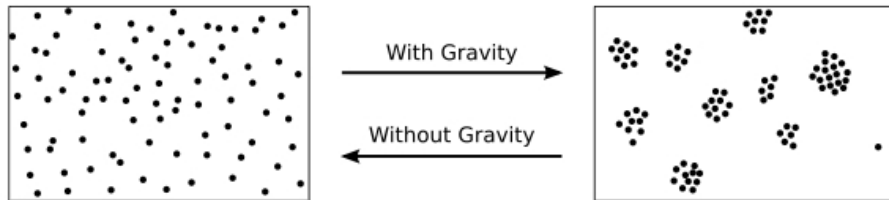


Figure 1: Effect of switching gravity on or off.

Weakening Gravity

What are the ways in which we could obtain homogeneous initial conditions? One option is to plainly declare that we have no business imposing any naturalness requirements on initial conditions. Initial conditions after all can be freely chosen and we can simply choose ones that do the trick and leave it at that. This may well turn out to be the answer but I, and many others, find it dissatisfying. What we would prefer is a *dynamical* origin to the arrow of time. Yet, this seems impossible. If one wants to arrive at a state with low entropy dynamically then, by the second law, the only possibility seems to be to start with a state of even lower entropy! This is of course the same trap that inflation falls into.

Let us set aside the second law for a moment. Instead of thinking about entropy, suppose we focus on arriving at a homogeneous distribution of matter. How can we obtain a homogeneous state without assuming even greater homogeneity? The key idea [1] is that the presence or absence of gravity changes the equilibrium configuration of matter. Whereas in the presence of gravity a homogeneous distribution of matter would be considered to have low entropy (clumps being preferred), when gravity is switched off, this is of course a *high* entropy configuration.

We can now make out the contours of a solution: if gravity were somehow switched off, then, for generic choices of the initial state, the universe would evolve naturally to the equilibrium configuration. In the absence of gravity, this would of course be a homogeneous configuration. If we now switch on gravitational interactions, the homogeneous state would start to clump, to evolve into our observed present universe. But since the homogeneous configuration is an attractor for generic initial states in the absence of gravity, we would have succeeded in solving the arrow of time problem dynamically, without invoking very special initial conditions.

Several specific realizations of this idea are possible. One could consider a scenario, in Brans-Dicke theory for example, in which Newton's constant was smaller in the early universe [1]. Another possibility is to have a time-dependent effective mass for the graviton, mediated again through interaction with a scalar field. If the inverse mass of the graviton were smaller than the Jeans' length, gravity would effectively be a short-range force and the preferred

configuration of matter would be a homogeneous distribution. Finally, a string theory-based model [4] has been proposed in which only tachyons exist in the early universe; the graviton effectively disappears from the spectrum. There may also be other realizations of this general idea. In all of them, the central consequence of switching off gravity is to make a homogeneous distribution the natural endpoint of generic initial conditions.

Nevertheless, a nagging feeling may persist that if we take everything properly into account, we might find that we have sneaked in some special conditions along the way. When we switch on gravity, haven't we gone from high entropy to low entropy? What happened to the second law? One way out might be to make the entropy ill-defined. (This is similar in spirit to the model of [5] in which the maximum entropy is unbounded so that all states have "low" entropy compared to S_{max} .) More concretely, when we say that entropy is maximized, it is always subject to holding fixed some thermodynamic parameters, usually the energy and volume. When these are not fixed, there is of course no upper bound on the entropy. Now, in general relativity energy is a subtle quantity; it is only well-defined if the spacetime satisfies certain global properties. It may well be that our own cosmological spacetime does not have the requisite properties to define energy. If so, then the total or maximal entropy might also be ill-defined. It would be ironic if the problems with defining energy in cosmology are the ones that help us to define an arrow of time for our own universe.

References

- [1] B. Greene, K. Hinterbichler, S. Judes, and M. Parikh, "Weak gravity, smooth initial conditions, and the arrow of time," to appear.
- [2] S. Hollands and R. Wald, "An alternative to inflation," *Gen. Rel. Grav.* **34** (2002) 2043.
- [3] T. Padmanabhan, "Statistical mechanics of gravitating systems," *Phys. Repts.* **188** (1990) 285.
- [4] J. McGreevy and E. Silverstein, "The tachyon at the end of the universe," *JHEP* **0508** (2005) 090.
- [5] S. Carroll and J. Chen, "Spontaneous inflation and the origin of the arrow of time," [hep-th/0410270](#).