BEFORE THE BIG BANG: AN OUTRAGEOUS NEW PERSPECTIVE AND ITS IMPLICATIONS FOR PARTICLE PHYSICS

Roger Penrose
Mathematical Institute, 24-29 St Giles', Oxford OX1 3LB, U.K.

THE BASIC CONUNDRUM

Proposals for describing the initial state of the universe hardly ever address a certain fundamental conundrum [1] — yet this is a conundrum whose significance is, in a certain sense, obvious. The issue arises from one of the most fundamental principles of physics: the *Second Law of thermodynamics*. According to the Second Law, roughly speaking, the entropy of the universe increases with time, where the term "entropy" refers to an appropriate measure of *disorder* or lack of "specialness" of the state of the universe. Since the entropy increases in the future direction of time, it must decrease in the past time-direction. Accordingly, the initial state of the universe must be the most special of all, so any proposal for the actual nature of this initial state must account for its extreme specialness. Proposals have been put forward from time to time (such as in various forms of "inflationary cosmology" and the previously popular "chaotic cosmology") in which it is suggested that the initial state of the universe ought to have been in some sense "random", and various physical processes are invoked in order to provide mechanisms whereby the universe might be driven into the special state in which it appears actually to have been in, at slightly later stages. But "random" means "non-special" in the extreme; hence the conundrum just referred to.

Sometimes theorists have tried to find an explanation via the fact that the early universe was very "small", this smallness perhaps allowing only a tiny number of alternative initial states, or perhaps they try to take refuge in the *anthropic principle*, which would be a selection principle in favour of certain special initial states that allow the eventual evolution of intelligent life. Neither of these suggested explanations gets close to resolving the issue, however. It may be seen that, with time-symmetrical dynamical laws, the mere smallness of the early universe does not provide a restriction on its degrees of freedom. For we may contemplate a universe model in the final stages of collapse. It must do *something*, in accordance with its dynamical laws, and we expect it to collapse to some sort of complicated space-time singularity, a singularity encompassing as many degrees of freedom as were already present in its earlier non-singular collapsing phase. Time-reversing this situation, we see that an initial singular state could also contain as many degrees of freedom as such a collapsing one. But in our actual universe, almost all of those degrees of freedom were somehow not activated.

What about the anthropic principle? Again, this is virtually no help to us whatever in resolving our conundrum. It is normally assumed that life had to arise via complicated evolutionary processes, and these processes required particular conditions, and particular physical laws, including the Second Law. The Second Law was certainly a crucial part of evolution, in the way that our particular form of life actually came about. But the very action of this Second Law tells us that however special the universe may be now, with life existing in it now, it must have been far *more* special at an earlier stage in which life was not present. From the purely anthropic point of view, this earlier far more special phase was not needed; it would have been much more likely that our present "improbable" stage came about simply by chance, rather than coming about via an earlier even more improbable stage. When the Second Law is a crucial component, there is always a far more probable set of initial conditions that would lead to this same state of affairs, namely one in which the Second Law was *violated* prior to the situation now!

As another aspect of this same issue, we may think of the vastness of our actual universe, most of which had no actual bearing on our existence. Though very special initial conditions were indeed required for our existence in our particular spatial location, we did not actually need these same special conditions at distant places in the universe. Yet as we look out at the universe, we see the same kind of conditions, acting according to the same Second Law of thermodynamics, no matter how far out we look. If we take the view that the Second Law was introduced in our vicinity merely for our own benefit, then we are left with no explanation for the extravagance of this same Second Law having to be invoked uniformly throughout the universe, as it appears to be as far as our powerful instruments are able to probe.

THE ENORMITY OF THE SPECIALNESS

In order to stress the extraordinary scale of this problem, and the intrinsic implausibility of explanations of this kind, it is helpful to enter a little more precisely into the definition of entropy, and to estimate the entropy magnitudes that we have to contend with. Boltzmann provided us with a beautiful formula for the entropy S of a system:

$$S = k \log V.$$

Here k is Boltzmann's constant and V is the volume of a certain region in the total phase space $\boldsymbol{P}$ of the system under consideration. We are assuming $\boldsymbol{P}$ to be "coarse-grained" into sub-regions, each sub-region representing states that are deemed to be indistinguishable with regard to any reasonable macroscopic parameter. (There is clearly an element of arbitrariness or subjectivity, here, as to which parameters are to be regarded as macroscopically discernible and which are deemed to be effectively "unmeasurable". In practice, there is a considerable robustness with regard to this arbitrariness, and it is reasonable to disregard this issue in the present discussion.) Any particular state of the system under consideration will be specified by some point x of $\boldsymbol{P}$, and the quantity V is then the volume of the particular sub-region of $\boldsymbol{P}$ which contains x.

With regard to future time-evolution of the system, the Second Law can be understood as the fact that, as the system evolves, the point x moves within $\boldsymbol{P}$ so that with overwhelming probability it enters sub-regions of successively larger and larger volume V. This arises from the fact that, in practice, the sub-regions differ stupendously in size. The logarithm in Boltzmann's formula helps here (as does the smallness of k, in ordinary units), because there need only be a modest increase in S when x moves from one sub-region into a neighbouring one of stupendously larger volume. But this is only the easy half of our understanding of the Second Law. The difficult half is to understand why, when we *reverse time*, x enters successively *tinier* sub-regions of $\boldsymbol{P}$. It does this because it has ultimately to reach the exceptionally tiny region $\boldsymbol{B}$ which represents the Big Bang itself. The difficult half of the Second Law involves an understanding of why the universe had to start off in such an extraordinarily special state. And to understand how special the Big Bang actually was, we need to compare the volume of $\boldsymbol{B}$ with that of the entire phase space $\boldsymbol{P}$.

One point of concern is the fact that the entire volume might be infinite, as it certainly would be in the case of a spatially infinite universe. This issue, while of relevance, is not of major importance for our considerations here. There is also the issue of how we get a finite phase-space volume when some of the parameters would be describing continuous fields. I shall evade this latter issue by assuming that it is dealt with by quantum mechanics, where for a finite universe of bounded energy content we may assume only finitely many quantum states.

To deal with a spatially infinite universe, I shall assume that we need consider only, say, that comoving portion of the universe that intersects our past light cone. This contains something of the order of 10^{80} baryons. To obtain a lower bound for the volume of $\boldsymbol{P}$, for this situation, we can consider the entropy that arises when this number of baryons is collapsed into a black hole. For this, we use the Bekenstein–Hawking entropy formula $S_{BH}=8\pi^2kGm^2/hc$ for a spherical black hole of mass m and find a value of the order of 10^{123}. If we collapsed the dark matter also into this black hole, we would get a considerably larger entropy (and, for a continually expanding universe, we should consider even larger values than this), but this value represents a usable lower bound. Recalling the logarithm in Boltzmann's formula (a natural logarithm, but that is of no concern), we get that the volume of $\boldsymbol{P}$ is greater than that of $\boldsymbol{B}$ by a factor that exceeds

$$10^{10^{123}}.$$

This gives us some idea of the enormity of the precision in the Big Bang!

THE GEOMETRIC NATURE OF THE SPECIALNESS

A seeming paradox arises from the fact that our best evidence for the very existence of the Big Bang arises from observations of the microwave background radiation—frequently referred to as the "flash of the Big Bang", greatly cooled down to its present value of ~2.7K. The intensity of this radiation, as a function of frequency, matches the Planck radiation formula extraordinarily closely, giving us impressive evidence of an early universe state with matter in *thermal equilibrium*. But thermal equilibrium is represented, in phase space $\boldsymbol{P}$, as the coarse-graining sub-region of largest volume (so large that it normally exceeds all others put together). This corresponds to *maximum* entropy, so we reasonably ask: how can this be consistent with the Second Law, according to which the universe started with a very *tiny* entropy?

The answer lies in the fact that the high entropy of the microwave background refers only to the *matter* content of the universe and not to the *gravitation field*, as would be encoded in its space-time geometry in accordance with Einstein's general relativity. What we find, in the early universe, is an extraordinary *uniformity*, and this can be interpreted as the gravitational degrees of freedom that are potentially available to the universe being not excited at all. As time progresses, the entropy rises as the initially uniform distribution of matter begins to clump, as the gravitational degrees of freedom begin to be taken up. This allows stars to be formed, which become much hotter than their surroundings (a thermal imbalance that all life on Earth depends upon), and finally this gravitational clumping leads to the presence of black holes (particularly the huge ones in galactic centres), which represent an enormous increase in entropy.

Although, in general, there is no clear geometric measure of the entropy in a gravitational field in general relativity, we can at least provide proposals for the *non-activation* of gravitational degrees of freedom at the Big Bang. I have referred to such a proposal as the *Weyl Curvature Hypothesis* (WCH) [2]. In Einstein's theory the *Ricci curvature* R_{ab} is directly determined by the gravitational sources, via the energy-momentum tensor of matter (analogue of the charge-current vector J^a in Maxwell's electromagnetic theory) and the remaining part of the space-time Riemann curvature, namely the *Weyl curvature* C_{abcd}, describes gravitational degrees of freedom (analogue of the field tensor F_{ab} of Maxwell's theory). WCH—which is a time-asymmetrical hypothesis—asserts that *initial* space-time singularities must be constrained to have C_{abcd}=0 (in some appropriate sense), whereas final space-time singularities (as occur inside black holes) are unconstrained.

What appears to be the most satisfactory form of WCH has been studied extensively by Paul Tod [3]. This proposes that an initial space-time singularity can always be represented as a smooth *past boundary* to the *conformal geometry* of space-time. In conformal geometry, we consider the space-time structure that is invariant under rescalings of the metric

$$g_{ab} \;\rightarrow\; \hat{g}_{ab} \;=\; \Omega^2\, g_{ab}$$

where Ω is taken to be a smooth positive scalar field on space-time. Another way of specifying the conformal geometry of space-time is simply to take the family of *null cones* (in the tangent spaces of the space-time points) as defining the geometry. It may be noted that the conformal structure contains 9 out of the 10 components of the metric, the overall scale providing, in effect, the 10th.

Tod's formulation of WCH is the hypothesis that we can adjoin a (past-spacelike) hypersurface boundary to space-time in which the conformal geometry can be mathematically extended smoothly through it, to the past side of this boundary. This amounts to "stretching" the metric by a conformal factor Ω which becomes *infinite* at the Big Bang singularity, so that we get a smooth metric $\hat{g}_{ab}$ which actually extends across this boundary.

CONFORMAL CYCLIC COSMOLOGY

So far, we regard the conformal "space-time" prior to the Big Bang as a mathematical fiction, introduced solely in order to formulate WCH in a mathematically neat way. However, my "outrageous" proposal [4] is to take this mathematical fiction seriously as something *physically real*. But what "physical reality" can we consistently attach to this space-time occurring "before the Big Bang"? As a clue to this possibility, we should consider the nature of the physics that is presumed to be taking place just *after* the Big Bang. (I am going to ignore the possibility of inflation here, and assume that such exponential expansion did not actually take place after the Big Bang. The issue of inflation, in relation to this scheme, is considered in the section on physical implications, below.) As we approach the Big Bang, moving back in time, we expect to find temperatures that are increasingly great. And the greater the temperature, the more irrelevant the rest masses of the particles involved will become, so these particles are effectively massless near the Big Bang. Now, massless particles (of whatever spin) satisfy conformally invariant equations [5]. I am going to suppose that the *interactions* between these massless entities are also described by conformally invariant equations. (This seems to be consistent with current understanding of particle physics.) With such conformal invariance holding in the very early universe, the universe has no way of "building a clock". So it loses track of the scaling which determines the full space-time metric, while retaining its conformal geometry.

We may apply considerations of this kind also to the distant future of the universe. If we assume that in the very remote future, conformally invariant equations again govern the universe's contents, then we can apply the same mathematical trick as before, but now in the reverse sense that we look for a boundary at which the conformal factor Ω becomes *zero*, rather than infinite. This amounts to using a metric, such as $\hat{g}_{ab}$ above, in which the future infinity is "squashed down" to be a finite boundary to space-time, which is conformally regular in the sense that the space-time can be mathematically extended across this future boundary as a smooth conformal manifold [5]. If we also assume that there is a positive cosmological constant present, as current observations appear to point strongly towards, then we find that this future conformal boundary is *spacelike*.

There is, however, a crucial difference between the use of a conformal boundary to study the future asymptotics of a space-time and Tod's use of a conformal boundary to treat the Big Bang. For in the latter case the very validity of this trick provides a formulation of WCH, whereas it the future situation of an expanding universe with conformally invariant contents, the validity of this procedure is more-or-less *automatic* [5]. Physically, we may think that again in the very remote future, the universe "forgets" time in the sense that there is no way to build a clock with just conformally invariant material. This is related to the fact that massless particles, in relativity theory, do not experience any passage of time. We might even say that to a massless particle, "eternity is no big deal". So the future boundary, to such an entity is just like anywhere else. With conformal invariance both in the remote future and at the Big-Bang origin, we can try to argue that the two situations are *physically identical*, so the remote future of one phase of the universe becomes the Big Bang of the next. This suggestion is my "outrageous" *conformal cyclic cosmology*" (CCC) [4].

PHYSICAL IMPLICATIONS

There are certain important assumptions involved in CCC, in order that only conformally invariant entities survive to eternity . One of these is that black holes will all eventually evaporate away and disappear. This evaporation is a consequence of Stephen Hawking's quantum considerations, and these are now normally accepted. There is, however, the issue (connected with the so-called "information paradox") of whether they would actually ultimately disappear or leave some form of "remnant". I am here taking the more conventional view that they would indeed disappear in a final (cosmologically very mild) explosion. More serious, for CCC, is how to get rid of massive fermions and massive charged particles. It is not too unconventional to assume that protons will ultimately decay, or even that there is one variety of neutrino that is massless, but the real problem lies with electrons. A good many of them will annihilate with positively charged particles, but there will be a relatively small number of "stray" charged particles which become trapped in their ultimate event horizons, being unable to come in contact with other particles of opposite charge. There are various possible ways out of this, none of which is part of conventional particle physics. One possibility is that electric charge is not exactly conserved, so that within the span of eternity, electric charge would eventually disappear. A much more satisfying possibility, from my own perspective, is that the electron's mass will eventually decay away—and, again, there is all of eternity for this to happen, so the possibility may not be too outrageous to contemplate.

This last possibility is tied up with the issue of the strength of the gravitational interaction, which I have postponed in my discussion here. In the background of conformal geometry, the strength of gravity may be considered as being infinitely large at the Big Bang (which is, in a sense, why the gravitational degrees of freedom must initially be set to zero), and this strength gets smaller as time progresses, eventually reducing to zero at the final boundary. To express all this in a satisfactory mathematical framework for CCC, we need to reformulate general relativity in a conformally invariant way. This can indeed be done. We take advantage of the fact that the Weyl tensor $C_{abc}{}^{d}$ is conformally invariant, and provides a precise measure of the conformal curvature of space-time. We can define the gravitational "spin-2 field" $K_{abc}{}^{d}$ to be described by $C_{abc}{}^{d}$ with respect to the original space-time metric g_{ab}, but when we pass to the conformally related metric $\hat{g}_{ab}=\Omega^2 g_{ab}$ we find that, curiously, $K_{abc}{}^{d}$ picks up a factor of Ω^{-1}, which $C_{abc}{}^{d}$ does not [5].

This has the implication that gravitational radiation (described by $K_{abc}{}^{d}$) actually survives at the future boundary (whereas $C_{abc}{}^{d}$ vanishes there) and its presence shows up as a non-zero normal derivative of $C_{abc}{}^{d}$ at the boundary. This gives rise to density fluctuations at the Big bang, and possibly primordial gravitational radiation. The details of all this have yet to be worked out, but, in principle at least, there should be clear-cut predictions which should be observable. One important issue is how this compares with the detailed observations of temperature variations in the microwave background and the near scale-invariance of the initial density fluctuations. This scale invariance is normally taken as a success of inflationary theory. It will be interesting to see whether CCC leads to a similar implication with regard to these fluctuations, as it also involves an exponential expansion, though this occurs *before* the Big Bang in CCC, rather than afterwards.

REFERENCES

[1] See, for example, Penrose, R. (2004) *The Road to Reality: A complete guide to the Laws of the Universe* (Jonathan Cape, London); Chapter 27.

[2] Penrose, R. (1979) Singularities and Time-Asymmetry, in *General Relativiy: an Einstein Centenary,* eds. S.W. Hawking and W. Israel (Cambridge University Press).

[3] Tod K.P. (2003) *Class. Quantum Grav.* **20** 521-534.

[4] Penrose, R (2005) http://www.newton.cam.ac.uk/webseminars/pg+ws/2005/gmr/ gmrw04/1107/ penrose/. Related ideas, but dependent upon extra space dimensions, were put forward earlier: Veneziano, G. (1998), see A Simple/Short Introduction to Pre-Big-Bang Physics/Cosmology, hep-th/9802057 v2 in the LANL Archive; also Steinhardt, P.J. and Turok N. (2001) A Cyclic Model of the Universe http://xxx.lanl.gov/PS_cache/hep-th/pdf/0111/0111030.pdf

[5] Penrose, R. (1965) Zero rest-mass fields including gravitation: asymptotic behaviour, *Proc. Roy. Soc. London*, **A284**, 159-203.