Proc. R. Soc. Lond. A. **338**, 439–446 (1974)
Printed in Great Britain

Cosmological models and the Large Numbers hypothesis

By P. A. M. Dirac, O.M., F.R.S.
Florida State University, Tallahassee, Florida

(*Received* 3 *December* 1973)

The Large Numbers hypothesis asserts that all the large dimensionless numbers occurring in Nature are connected with the present epoch, expressed in atomic units, and thus vary with time. It requires that the gravitational constant G shall vary, and also that there shall be continuous creation of matter.

The consistent following out of the hypothesis leads to the possibility of only two cosmological models. One of them, which occurs if one assumes that the continuous creation is a multiplication of existing matter, is Einstein's cylindrical closed Universe. The other, which occurs if one assumes the continuous creation takes place uniformly through the whole of space, involves an approximately flat Minkowski space with a point of origin where the Big Bang occurred.

1. The Large Numbers hypothesis

The hypothesis has been discussed by the author (1973*a*) as the revival of an old idea. If one expresses the age of the Universe in terms of a unit of time provided by atomic constants, say e^2/mc^3, one gets a large dimensionless number t, which is somewhere around 10^{39}. It characterizes the present epoch in a natural way, independent of man-made standards.

The fundamental data provided by astronomy and atomic physics enable one to obtain other large dimensionless numbers. The Large Numbers hypothesis asserts that such numbers are connected with t, and therefore vary as t varies. A number which is roughly $(10^{39})^n$ must vary in proportion to t^n. The reason for believing the hypothesis is that without it one does not see how these large numbers could ever be explained.

The hypothesis is not easy to fit in with cosmological theories. It puts a *severe restriction* on the permissible models of the Universe. A popular model involves the Universe expanding to a certain maximum size and then contracting again. The time of maximum expansion, expressed in atomic units, will provide a large dimensionless number which does not depend on the present epoch t. It is just such numbers that the hypothesis rules out. A model with a maximum size for the Universe is thus not permitted.

Various models have been proposed by Friedman (1922, 1924) and others, involving non-static spherically symmetric solutions of the Einstein equations. They each refer to some characteristic epoch appearing in the equations, governing the expansion. This epoch, expressed in atomic units, again gives a large number independent of t, the present epoch, and thus a constant. Hence, the model is not allowed.

The only surviving models are those whose equations do not refer to a particular constant epoch. They are essentially static universes.

2. Physical consequences of the hypothesis

The electric force between the electron and the proton in a hydrogen atom is e^2/r^2. The gravitational force between them is $Gm_p m_e/r^2$. Their ratio is the dimensionless number $e^2/Gm_p m_e$. Its value is about 2×10^{39}. According to the hypothesis it should be increasing proportional to t. Thus G, expressed in atomic units, should be decreasing proportional to t^{-1}.

This is a physical effect which should show up with sufficiently accurate measurements. The most hopeful chance of observing it is with Shapiro's (1968) radar measurements of the distances of the planets. If there is a secular variation of these distances which cannot be explained in any other way, it will provide evidence for a variation of G. Shapiro's observations are sufficiently accurate for an effect of the expected order of magnitude to show up in a few years' time.

There is also a possibility of observing the variation of G directly from laboratory experiments. Measurements of G are being carried out by Beams (1971) by an improved technique. They are not yet very accurate, but variations of G can be observed with greater accuracy than G itself and it may be that the apparatus can be improved sufficiently to show up the desired effect.

If one estimates the total number of nucleons in the Universe (or the total number in the galaxies with a speed of recession of less than $\frac{1}{2}c$, if one considers the Universe to be infinite) one gets a number somewhere around 10^{78}. According to the hypothesis, this must be increasing proportional to t^2. It follows that new nucleons must continually be created.

Some time ago the Steady State model of the Universe was very popular. This required continuous creation of matter to balance the matter that was moving away from us with the recession of the galaxies. The Steady State model of course requires that G shall be constant, and is thus not consistent with our basic hypothesis. The theory being developed here works with the Big Bang model, but it also requires continuous creation of matter.

The continuous creation that we are forced to adopt is a new physical process, a kind of radioactivity, which is quite different from all the observed radioactivity. We must face the question of where this new matter is created. There are two alternative assumptions that one might make.

One might assume that nucleons are created uniformly throughout space, and thus mainly in intergalactic space. We may call this *additive creation*.

One might assume that new matter is created where it already exists, in proportion to the amount existing there. Presumably the new matter consists of the same kind of atoms as those already existing. We may call this *multiplicative creation*.

3. The two metrics

Einstein's theory of gravitation demands that G shall be a constant. In fact, with a natural choice of units $G = 1$. Now Einstein's theory is very successful in accounting for observations and we do not wish to abandon it. We therefore have to face the problem of how to modify it to make it agree with a slowly varying G while not spoiling its successes.

One can achieve this by supposing that the metric $\mathrm{d}s_{\mathrm{E}}$ occurring in the Einstein field equations is not the same as the metric $\mathrm{d}s_{\mathrm{A}}$ measured by atomic apparatus. With both metrics we take the velocity of light $c = 1$. All distances determined by atoms, e.g. the wavelengths of spectral lines and lattice spacings in crystals, refer to $\mathrm{d}s_{\mathrm{A}}$, so all laboratory measurements of distances and times give $\mathrm{d}s_{\mathrm{A}}$. One cannot measure $\mathrm{d}s_{\mathrm{E}}$ directly. It comes into play only in equations of motion. For example, calculations of the motions of planets involve $\mathrm{d}s_{\mathrm{E}}$. If one measures the distances of the planets with laboratory apparatus, as Shapiro is doing, one gets directly the ratio of the two ds.

Let us determine the connexion between the two ds. Take as an example the motion of the Earth around the Sun, in Newtonian approximation. The basic equation is

$$GM = v^2 r,$$

where M is the mass of the Sun, r is the radius of the Earth's orbit and v is the Earth's velocity. The formula applies both in Einstein units, when we may write it

$$G_{\mathrm{E}} M_{\mathrm{E}} = v_{\mathrm{E}}^2 r_{\mathrm{E}},$$

and in atomic units, when we may write it

$$G_{\mathrm{A}} M_{\mathrm{A}} = v_{\mathrm{A}}^2 r_{\mathrm{A}}.$$

Referred to Einstein units, all the quantities G_{E}, M_{E}, v_{E}, r_{E} are constants. Now v is dimensionless, it is just a certain fraction of the velocity of light, so $v_{\mathrm{A}} = v_{\mathrm{E}} =$ constant. From our previous discussion $G_{\mathrm{A}} :: t^{-1}$. The way M_{A} behaves depends on which assumption about creation we adopt. With additive creation, the number of nucleons in the Sun is constant, so M_{A} is constant. With multiplicative creation the number of nucleons $:: t^2$ so $M_{\mathrm{A}} :: t^2$. Thus with additive creation we get $r_{\mathrm{A}} :: t^{-1}$ and with multiplicative creation $r_{\mathrm{A}} :: t$. These r_{A} are to be compared with a constant r_{E}. The general result is

$$\mathrm{d}s_{\mathrm{A}} = t^{-1}\,\mathrm{d}s_{\mathrm{E}} \quad \text{additive creation,} \quad \text{(I)}$$

$$\mathrm{d}s_{\mathrm{A}} = t\,\mathrm{d}s_{\mathrm{E}} \quad \text{multiplicative creation.} \quad \text{(II)}$$

With additive creation, the Earth is approaching the Sun (in atomic units) and the whole Solar System is contracting. With multiplicative creation, the Earth is receding from the Sun and the whole Solar System is expanding. These effects are cosmological and are to be superposed on other effects arising from known physical causes. Shapiro's observations should show them up if they exist, and should enable one to distinguish between the two kinds of creation.

Let us adopt the Einstein metric and proceed to examine the possible cosmological models. We can then use the Einstein field equations.

If we take the field equations with the cosmological constant λ, then the theory involves a large distance $R = \lambda^{-\frac{1}{2}}$, of the order of the radius of the Universe. Let δs_E be an atomic distance, such as the wavelength of a certain spectral line, expressed in the Einstein metric. Then $R/\delta s_E$ is dimensionless, and it is a large number of the order of 10^{39}. Thus it increases proportional to t. Now R is a constant, so $\delta s_E :: t^{-1}$. If the atomic distance is expressed in atomic units, to give ds_A, it is, of course, constant. It follows that we have case II, multiplicative creation.

We can infer that if we have the cosmological term in the Einstein equations, we must have multiplicative creation. Additive creation can occur only if $\lambda = 0$.

4. Multiplicative creation

Let us consider further the case of multiplicative creation. We may introduce the Einstein epoch τ,

$$\tau = \int ds_E, \tag{1}$$

taken along the world-line of a galaxy. From (II), putting $ds_A = dt$,

$$\tau = \int t^{-1}\,dt = \ln t. \tag{2}$$

The variable τ is the dynamical time, as distinct from the atomic time t. The Big Bang occurred at $t = 0$, corresponding to $\tau = -\infty$, so referred to dynamical time the Universe has always existed.

The idea of two time variables related in this way was first introduced by Milne (1937), but his theory was based on quite different assumptions from the present one, and it is just a coincidence that they both give formula (2).

The definition of energy in the Einstein theory has an ambiguity connected with the coordinate system, but this does not come into play in cosmological models. We must then have conservation of energy or of mass (if we ignore the possibility of an appreciable pressure working on the expansion of the Universe). We shall then have the Einstein equations applying with the metric ds_E and with a suitable unit of mass. This unit must be such that the mass of a body such as the Sun is constant. The mass of a nucleon must then be proportional to t^{-2}. All atomic particles must have their masses varying in this way. It just compensates the multiplicative creation to make the masses of classical bodies constant.

One can easily work out how all the atomic constants vary when referred to Einstein units. We had in § 2

$$e^2/Gm^2 :: t.$$

In Einstein units G is constant and $m :: t^{-2}$. Hence, $e :: t^{-\frac{3}{2}}$. We then find $h :: t^{-3}$.

We saw in § 1 that we must have a static model. The only static model with positive mass is Einstein's cylindrical model, which we are thus forced to adopt with multiplicative creation.

We now have a picture in which, referred to Einstein units, the galaxies are not receding, but keep approximately at a constant distance. To understand how the red-shift arises with this picture we must take into account that an atomic clock, marking out units $\Delta t = 1$, will mark out units $\Delta\tau = t^{-1}$. With increasing t these units get continually smaller, so the atomic clock is continually speeding up.

The light coming from a distant galaxy was emitted in the past when atomic clocks were slower. The wavelength of the emitted light referred to these slow atomic clocks. As the light travels to us the wavelength remains constant in Einstein units. When it arrives here it is referred to the present atomic clocks and the wavelength appears longer. The increase is in the proportion t/t_e, where t_e is the time of emission. Thus, the red-shift is $t/t_e - 1$.

5. Additive creation

Let us now examine the alternative kind of creation. We now have matter, presumably hydrogen atoms, created uniformly throughout space. This will give violation of conservation of mass whatever units we use. The only way in which we can preserve the Einstein equations, which demand conservation of mass, is to suppose that, together with the H atoms, a uniform distribution of negative mass is created, so as to make the total density of created matter zero. In order not to have violent disagreement with observation we must suppose that the negative mass is not observable and so is not quantized, like the H atoms. It must not interact with other matter, except gravitationally, and must have no physical effects at all, apart from producing a curvature of space. The H atoms condense into nebulae and stars and form the matter that we observe. The negative mass remains uniform and unobservable.

The total density of matter is zero apart from local irregularities arising from condensations of the H atoms. If we smooth out these irregularities we get a model in which space–time, referred to the Einstein metric, is flat. It is thus just Minkowski space and the Einstein metric becomes the Minkowski metric.

There is one special point, the origin O, where the Big Bang occurred. The physical world lies within the future light cone from O. The world-lines of the galaxies are the straight lines through O lying within the future light cone.

The Einstein epoch τ at a physical point P is the Minkowski distance from P to O. Formula (1) still applies and must now be used in conjuction with I, so

$$\tau = \int \mathrm{d}s_{\mathrm{E}} = \int t\,\mathrm{d}t = \tfrac{1}{2}t^2. \tag{3}$$

The three-dimensional world at a given epoch consists of all those points P with a given τ-value. It is of infinite extent and has a negative curvature.

The total number of nucleons within that part of the world whose speed of recession is $< \frac{1}{2}$ is proportional to t^2 or to τ. The volume of this part of the world is $:: \tau^3$. The number per unit volume is thus $:: \tau^{-2}$. With additive creation the mass of a nucleon in Einstein units in constant, as is shown by the constancy of M_{E} in § 3.

Thus the density of the nucleons :: τ^{-2}. The density of the continuous distribution of negative mass must therefore also be proportional to τ^{-2}.

The Einstein field equation is

$$R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R = 8\pi(T^{\mu\nu} - U^{\mu\nu}), \tag{4}$$

where $T^{\mu\nu}$ is the material energy tensor for the ordinary physical matter and $-U^{\mu\nu}$ is the corresponding tensor for the uniform negative matter. We have

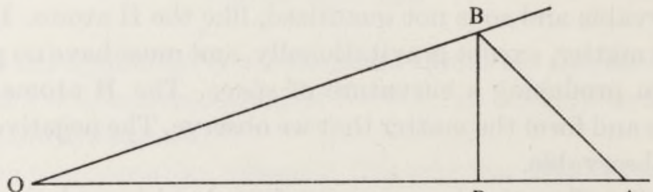
$$U^{\mu\nu} = k\tau^{-2}v^{\mu}v^{\nu},$$

where v^{μ} is the velocity vector corresponding to the recession speed for the point concerned and k is a constant. Thus we may rewrite equation (4)

$$R^{\mu\nu} - \tfrac{1}{2}g^{\mu\nu}R + k'\tau^{-2}v^{\mu}v^{\nu} = 8\pi T^{\mu\nu} \tag{5}$$

with another constant k'.

Because of the uniformity of the negative matter, the term $k'\tau^{-2}v^{\mu}v^{\nu}$ may be looked upon as an intrinsic property of space, like the cosmological term $\lambda g^{\mu\nu}$ that is introduced into other forms of the Einstein theory. Its effect on the motions of the planets is just as small as that of the usual cosmological term, so it is much too small to be observed. The only reason we have to put it into equation (5) is to preserve consistency with the Bianci identities when $T^{\mu\nu}$ involves continual creation.



Let us discuss the red-shift with this model. Let OA represent the world-line of our galaxy and let OB represent the world-line of another galaxy, and suppose the points A and B are such that light emitted from B can arrive at A. Let P be the point on OA such that PB is orthogonal to OA. Then PB = PA. If θ is the hyperbolic angle between OB and OA, we have

$$\mathrm{OP} = \mathrm{OB}\cosh\theta,$$
$$\mathrm{PA} = \mathrm{PB} = \mathrm{OB}\sinh\theta,$$

and hence

$$\mathrm{OA} = \mathrm{OB}\,\mathrm{e}^{\theta}.$$

Let τ_A and τ_B denote the epochs of A and B in Einstein units, so that they are equal to the distances OA, OB respectively. Then

$$\tau_A/\tau_B = \mathrm{e}^{\theta}.$$

If t_A, t_B are the epochs in atomic units, we get from (3)

$$t_A/t_B = \mathrm{e}^{\frac{1}{2}\theta}.$$

Suppose light is emitted from B with a wavelength $\lambda = \delta t_B$. It is received at A with the wavelength $\delta t_A = e^{\frac{1}{2}\theta}\,\delta t_B = e^{\frac{1}{2}\theta}\,\lambda$. The red-shift is thus $e^{\frac{1}{2}\theta} - 1$.

There would still be a red-shift with this model even if there were only one metric, the Minkowski metric. The magnitude of the red-shift would then be

$$\delta\tau_A/\delta\tau_B - 1 = e^{\theta} - 1.$$

For small θ it is just double the red-shift with the two-metric theory. We need the two-metric theory, of course, in order to have G varying with the epoch.

6. Comparison of the two models

Two cosmological models have been obtained, corresponding to the two alternative assumptions that one might make for the continuous creation of matter. They are both very simple models. The question arises, which should one prefer? A decisive answer may be obtained in a few years time from radar observations of the planets. In the meantime we may discuss them and try to assess which is the more likely.

Multiplicative creation requires that all forms of matter shall be multiplying, with the number of atoms increasing proportional to t^2. It is a little difficult to understand how this can take place in the case of a crystal. Presumably the new atoms must appear on the outside. The rate of multiplication is extremely small, so there is plenty of time for the new atoms to appear in the places most suitable for them. But during the course of geological ages the increase must be quite appreciable, and should be taken into account in any discussion of the formation of crystals in very old rocks. It might lead to insuperable difficulties.

Multiplicative creation requires also that the number of photons in a given beam of light shall increase, in order that the energy in Einstein units shall be conserved. This will cause the apparent brightness of a distant galaxy or quasar to be increased. The effect would not be directly observable because we do not know the absolute brightness at the time in the remote past when the light was emitted. One may make statistical assumptions and try to get some evidence for this effect. The question has been discussed in the author's paper (1973*b*).

Additive creation does not lead to such drastic departures from generally accepted ideas and does not face us with such difficult problems. It involves continual creation of intergalactic gas and presumably much of it remains in the gaseous form, but its density is much too small to be observable. This theory is not so likely to lead quickly to a clash with observations, as could happen with multiplicative creation.

The foregoing work is all founded on the Large Numbers hypothesis, in which I have great confidence. It also requires the assumption of two metrics, which is not so certain. The only reason for believing in the two metrics is that up to the present no alternative way of bringing in the Einstein theory has been thought of. But this situation could change.

References

Beams, J. W. 1971 *Physics Today*, May, p. 35.
Dirac, P. A. M. 1973*a* *Proc. R. Soc. Lond.* A **333**, 403.
Dirac, P. A. M. 1973*b* *Pont. Acad. Commentarii*, vol. 11 – N.46, 1.
Friedman, A. 1922 *Z. Phys.* **10**, 377.
Friedman, A. 1924 *Z. Phys.* **21**, 326.
Milne, E. A. 1937 *Kinematic relativity*, p. 36. Oxford University Press.
Shapiro, I. I. 1968 *Scient. Am.* **219**, 32.