A new basis for cosmology

By P. A. M. Dirac, F.R.S.
St John's College, Cambridge

(*Received* 29 *December* 1937)

1. Introduction

The modern study of cosmology is dominated by Hubble's observations of a shift to the red in the spectra of the spiral nebulae—the farthest parts of the universe—indicating that they are receding from us with velocities proportional to their distances from us. These observations show us, in the first place, that all the matter in a particular part of space has the same velocity (to a certain degree of accuracy) and suggest a model of the universe in which there is a *natural velocity* for the matter at any point, varying continuously from one point to a neighbouring point. Referred to a four-dimensional space-time picture, this natural velocity provides us with a *preferred time-axis* at each point, namely, the time-axis with respect to which the matter in the neighbourhood of the point is at rest. By measuring along this preferred time-axis we get an absolute measure of time, called the *epoch*.

Such ideas of a preferred time-axis and absolute time depart very much from the principles of both special and general relativity and lead one to expect that relativity will play only a subsidiary role in the subject of cosmology. This first point of view, which differs markedly from that of the early workers in this field, has been much emphasized recently by Milne.

We now feel the need for some new assumptions on which to build up a theory of cosmology. This need is partially satisfied by the assumptions, which Milne calls the Cosmological Principle, that, apart from local irregularities, the universe is everywhere uniform and has spherical symmetry (in three dimensions) about every point, for an observer moving with the natural velocity at that point. The assumption of uniformity is to be taken in its most general form, in which it requires that an observer on another nebula would see all general natural phenomena (for example, the red-shift of other nebulae) the same as we do. The observational evidence in favour of these assumptions is rather meagre, since only a small part of the universe is accessible to present-day telescopes, and this part shows quite large fluctuations from uniformity in the distribution of the spiral nebulae (Reynolds 1937). However, these assumptions are fairly plausible and have a great simplifying effect on the subject, and until there is more definite evidence of their inadequacy it does not seem worth while to try more complicated schemes.

Further assumptions are needed if we are to obtain definite answers to the main problems that suggest themselves in a study of cosmology. A possible further assumption is Milne's Dimensional Hypothesis (Walker 1936, p. 121), which requires that there shall be no constants with dimensions appearing in cosmological theory. This assumption is open to criticism, as there is no definite reason why the constants of atomic theory should not appear in cosmology—in fact, one would rather expect them to, since one would expect a closer connexion between the atom and the cosmos to show itself with a deeper understanding of Nature. An alternative assumption, which is free from this criticism and is more far-reaching, will be given in the next section and forms the main theme of the present paper.

2. The fundamental principle

The recession of the spiral nebulae with velocities proportional to their distances from us requires, if we assume these velocities to be roughly constant, that at a certain time in the distant past all the matter in the universe was confined within a very small volume. This time appears as a natural origin of time and provides us with a zero from which to measure the epoch of any event. Referred to this zero the present epoch, according to Hubble's data, is about 2×10^9 years.

Let us express this in terms of a unit of time fixed by the constants of atomic theory, say the unit e^2/mc^3. We then get the value 7×10^{38}. This turns out to be of the same order of magnitude as the ratio, γ say, of the electric to the gravitational force between an electron and a proton, namely, $2 \cdot 3 \times 10^{39}$. If we had used another atomic unit of time in which to express the present epoch, we should have obtained a value differing from the above one by at most a few powers of ten, which would not have affected the agreement with γ as to order of magnitude, when such large numbers as 10^{39} are concerned. The unit we chose, namely, e^2/mc^3, lies roughly in the geometric mean of all the units of time that we can construct simply from the atomic constants, namely (introducing also the proton mass M),

$$\frac{e^2}{mc^3}, \quad \frac{e^2}{Mc^3}, \quad \frac{h}{mc^2}, \quad \frac{h}{Mc^2}, \quad \frac{\hbar}{mc^2}, \quad \frac{\hbar}{Mc^2},$$

which are in the ratio

$$1, \quad 0 \cdot 0005, \quad 850, \quad 0 \cdot 46, \quad 137, \quad 0 \cdot 074.$$

We might have compared the epoch with the ratio of the electric to the gravitational force between two electrons, or between two protons, instead of between one proton and one electron, which would have given us a number 1800 times larger or smaller than γ respectively. In any case, however, we see there is a close agreement between the present epoch, expressed in atomic units, and the ratio of the gravitational to the electric force between two elementary particles. Such a coincidence we may presume is due to some deep connexion in Nature between cosmology and atomic theory. Thus we may expect it to hold not only at the present epoch, but for all time, so that, for example, in the distant future when the epoch is 10^{50}, we may expect γ will then be of the order 10^{50}. We are thus led to the result that a quantity γ, usually considered as a universal constant, must vary with the passage of great intervals of time.

A further study of cosmology leads to the appearance of other very large dimensionless numbers. These numbers all turn out to be of the order 10^{39} or sometimes 10^{78}. From a natural extension of the foregoing ideas we should expect all those numbers of the order 10^{39} to increase proportionally to the epoch, and all those of the order 10^{78} to increase proportionally to the square of the epoch. We have here a new principle appearing, that all the very large dimensionless numbers occurring in Nature are simple powers of the epoch, with coefficients of the order unity.

To get this principle in its most general form we should not make the assumption, which we made at the beginning of this section, that the velocity of recession of each spiral nebula is roughly constant. Without this assumption we can still talk about the epoch of an event, but we have no natural zero from which to measure it, so that only the difference of two epochs can enter into laws of nature. We must now use Hubble's constant, namely, the coefficient of proportionality between the red-shift and the distance, as one of the quantities from which very large dimensionless numbers are to be constructed (to replace our previous use of the present epoch as one of these numbers) and express our principle in the form: *Any two of the very large dimensionless numbers occurring in Nature are connected by a simple mathematical relation, in which the coefficients are of the order of magnitude unity.* If we can deduce from elementary considerations that some of these very large numbers vary with the epoch (as we shall find in the next section is the case), then they must all do so to preserve the mathematical relations between them.

This very general formulation of the principle does not enable one to draw exact conclusions with certainty. If, for example, we have two numbers a and b both of the order 10^{39}, we cannot with certainty conclude that

$$a = kb, \tag{1}$$

where k is a constant of order unity. Owing to our numerical coincidences being inaccurate by a few powers of ten (on account mainly of the uncertainty of which atomic units to use), we must allow k to differ from unity by a few powers of ten, and thus we may have instead of (1), for example,

$$a = kb \log b,$$

with a somewhat different k. In the present paper, for the sake of getting a definite theory, we shall ignore the possible occurrence of such logarithmic factors, or other similar factors that vary slowly with their arguments. We must then remember that the resulting theory will be valid only as a first approximation and may need amendment in the future by the insertion into the equations of functions that vary slowly with their arguments.

Essentially the same approximation is involved in the assumption, which is implied throughout this paper, that $\hbar c/e^2$ and M/m are constants. Future developments may require these quantities to vary slowly with the epoch.

3. The law of recession of the spiral nebulae

Let us take two neighbouring spiral nebulae and express the distance between them in terms of a unit of distance provided by the atomic constants, say the unit of time that we used in the preceding section multiplied by the velocity of light. The distance between the nebulae then becomes a dimensionless number, which will vary with the epoch in an unknown way, and which we call $f(t)$. On account of our assumptions of uniformity and spherical symmetry in § 1, $f(t)$ must be the same for any two neighbouring nebulae, except for an arbitrary constant factor. The determination of the form of $f(t)$, giving the law for the rate of recession of the spiral nebulae, is one of the main problems of cosmology.

Let us obtain Hubble's constant, the red-shift per unit distance, in terms of $f(t)$. The time taken by light to travel from one of our neighbouring spiral nebulae to the other is, since we are using units which make the velocity of light unity, just $f(t)$. If we consider two waves of light starting out from one of the nebulae at times δt apart, they will arrive at the other nebula at times $\delta t + f(t+\delta t) - f(t)$ apart, owing to the different times of transit for the two waves. Thus light which is emitted with the period δt will arrive with the period $\delta t + f(t+\delta t) - f(t)$, and the red-shift, namely, the change in period per unit period, is $\delta f(t)/\delta t = f'(t)$, so that Hubble's constant is $f'(t)/f(t)$. From Hubble's data this has the value at the present epoch $1 \cdot 4 \times 10^{-39}$.

We now bring into the argument the average density of matter ρ, which has a meaning from our assumption of uniformity. We take as unit of mass the mass of the proton or neutron, and we assume that matter is conserved when expressed in this unit. From this assumption of conservation we can infer that, owing to neighbouring nebulae separating from one another according to the law $f(t)$, the average density of matter will decrease according to the law

$$\rho \propto f(t)^{-3}. \tag{2}$$

The observed value for the average density of luminous matter is about 5×10^{-31} g. cm.$^{-3}$, which becomes, in our present units, about 7×10^{-45}. This value must be increased by a factor, which is very hard to estimate but is probably a few powers of ten, to get the total average density of all matter. Allowing for the inaccuracy caused by the uncertainty of which atomic units we ought to use, we see that the average density matter is of the same order of smallness as Hubble's constant. The reciprocals of these two quantities are two very large numbers, to which our fundamental principle is applicable, and which must therefore be connected like the a and b in equation (1). Thus

$$\rho = kf'(t)/f(t),$$

where k is a constant of the order of magnitude unity. Combining this with (2), we get

$$f(t)^{-3} \propto f'(t)/f(t),$$

and hence

$$f(t) \propto (t+a)^{\frac{1}{3}},$$

a being a constant of integration. By suitably choosing the zero from which we measure t, we may make this constant vanish and we then have

$$f(t) \propto t^{\frac{1}{3}}. \tag{3}$$

This gives us the law for the rate of recession of the spiral nebulae. The velocities of recession are not constant, as we provisionally assumed at the beginning of § 2, but vary proportionally to $f'(t)$ or $t^{-\frac{2}{3}}$. However, with this law of recession we still have a natural origin of time, namely, the zero of the t in (3), when all the nebulae were extremely close together. From (3) we have

$$t = \tfrac{1}{3} f(t)/f'(t), \tag{4}$$

showing that the present epoch is still of the order 10^{39}, and is, in fact, just a third of the value we gave it at the beginning of § 2 on the assumption of constant velocities of recession. The new value, equal to about 7×10^8 years, is rather small, being less than the age of the earth as usually calculated from data of radioactive decay, but this does not cause an inconsistency, since a thorough application of our present ideas would require us to have the rate of radioactive decay varying with the epoch and greater in the distant past than it is now.

Our deduction of (3) involves the assumption of conservation of mass when expressed in proton or neutron units, which means conservation of the number of protons and neutrons (apart from processes involving the transformation of the rest-energy of these particles to or from another form). There is no experimental justification for this assumption, since a spontaneous creation or annihilation of protons and neutrons sufficiently large to alter appreciably the law (3) would still be much too small to be detected in the laboratory. However, such a spontaneous creation or annihilation of matter is so difficult to fit in with our present theoretical ideas in physics as not to be worth considering, unless a definite need for it should appear, which has not happened so far, since we can build up a quite consistent theory of cosmology without it.

4. The curvature of space

Take all the points in space-time for which the epoch has some given value t. They will lie on a three-dimensional surface, which is everywhere orthogonal, in the sense of special relativity, to the natural time-axis. We call it the t-space. Our assumptions of uniformity and spherical symmetry in § 1 require that the t-space shall be everywhere uniform and spherically symmetrical. It follows that the t-space must be a *space of constant curvature*, the metric being provided by the atomic unit of distance that we had previously. (For a detailed study of this question, based on group theory, see Walker 1936.)

The curvature may be either positive, zero or negative. If it is positive, the t-space is finite and is like the three-dimensional surface of a sphere in four dimensions. If it is zero or negative, the t-space is infinite and is flat or hyperbolic respectively. Which of these three cases holds cannot, from considerations of continuity, depend on the value of t and must therefore be characteristic of space-time as a whole. To decide between these three cases forms another main problem of cosmology.

The case of positive curvature can easily be ruled out. In this case the total mass of the universe is finite and, expressed in the proton or neutron unit, is a very large number. From our assumption of conservation of mass, this large number must be independent of the epoch. We thus get a contradiction with our fundamental principle, according to which all very large numbers occurring in Nature must vary with the epoch, since some of them, namely, the reciprocals of Hubble's constant and of the average density, do.

The case of negative curvature can be ruled out in a similar but rather more complicated way. The total mass of the universe is not finite in this case, but we can work instead with the mass contained at time t within a sphere of radius equal to the radius of curvature of t-space. If we take a different epoch t_1, there will be a natural correspondence between points on the t_1-space and points on the original t-space (corresponding points being on the same nebula) and any element of distance in the t_1-space will equal the corresponding element of distance in the t-space multiplied by $f(t_1)/f(t)$. This factor being the same for all the elements of distance, it follows that the radius of curvature of the t_1-space must equal that of the t-space multiplied by this factor. The total mass contained within a sphere of radius equal to the radius of curvature must now be the same for the t_1-space as for the t-space. This mass, expressed in the proton or neutron unit, will again give us a constant number, which must be very large, in order that the curvature of t-space may be sufficiently small not to be in disagreement with observation, and which therefore contradicts our fundamental principle.

We are thus left with the case of zero curvature, or flat t-space, as the only one consistent with our fundamental principle and with conservation of mass. It should be remembered that the curvature we are here speaking about is the curvature of the three-dimensional space at one epoch and not the curvature of space-time as comes into general relativity.

5. The motion of a free particle

One other problem we shall concern ourselves with in this paper is the determination of the world-line of a particle that is moving freely under the action only of the gravitational field of the universe as a whole. We need something to replace Newton's first law of motion. If the particle is started off with the natural velocity of the place where it is situated, then, from our assumption of the spherical symmetry of the universe about any point, the particle cannot have an acceleration in any direction and Newton's law must hold for it. If, however, it is started off with a different velocity, then we cannot assert more than that its acceleration must lie in the plane in space-time containing its velocity vector and the natural velocity vector of the place. The magnitude of the acceleration may be any function of the velocity of the particle relative to the natural velocity and of the epoch.

Since general relativity explains so well local gravitational phenomena, we should expect it to have some applicability to the universe as a whole. We cannot, however, expect it to apply with respect to the metric provided by the atomic constants, since with this metric the "gravitational constant" is not constant but varies with the epoch. We have, in fact, from the discussion at the beginning of § 2, the ratio of the gravitational force to the electric force between electron and proton varying in inverse proportion to the epoch, and since, with our atomic units of time, distance and mass, the electric force between electron and proton at a constant distance apart is constant, the gravitational force between them must be inversely proportional to the epoch. Thus the gravitational constant will be inversely proportional to the epoch.

Let us try to set up a new system of units, whose ratios to the old units may vary with the epoch, so that with respect to the new units the gravitational constant does not vary with the epoch and general relativity may be expected to apply. We must not take a new unit of mass whose ratio to the old one varies with the epoch, as we should then have the mass of a proton or neutron varying with the epoch, and general relativity requires that the mass of an isolated particle shall remain constant. We must therefore change our units of distance and time, and must change both in the same ratio in order to keep the velocity of light unity. Since the dimensions of the gravitational constant are $(\text{distance})^3\,(\text{time})^{-2}\,(\text{mass})^{-1}$, we must take new units of distance and time equal to the old ones divided by the epoch, so that the new measure of a distance or time interval is equal to the old one multiplied by the epoch, to make the gravitational constant independent of the epoch.

We may now reasonably assume that, with the metric provided by the new measures of distance and time, general relativity holds and free particles move along geodesics. We thus have two measures of distance and time that are of importance, one for atomic phenomena and the other for ordinary mechanical phenomena included under general relativity. This situation is the same as Milne has with his two measures of time t and τ (Milne 1936, 1937*a*, 1937*b*), but the ratio of the two measures is just the inverse in our theory from what it is in Milne's.

6. The curvature of space-time

It is of interest to discuss the curvature of space-time, referred to the metric to which general relativity applies, and to determine to what stress-energy tensor it corresponds. This curvature is different, of course, from the curvature of t-space dealt with in § 4. Let us use stars to denote quantities measured in the new units, so that for an interval of time

$$\delta t^* = t\,\delta t$$

and thus

$$t^* = \tfrac{1}{2}t^2, \tag{5}$$

giving us the connexion between the new and the original measure of the epoch. The distance between neighbouring spiral nebulae, expressed in the new units, will vary with the epoch according to the law

$$f^*(t) = tf(t) \propto t^{\frac{4}{3}},$$

and hence

$$f^*(t^*) \propto t^{*\frac{2}{3}}. \tag{6}$$

We may now use Robertson's formulae (Robertson 1933, equations 3·2), according to which the curvature of our space-time must correspond to a uniform density ρ^* and a uniform hydrostatic pressure p^* given by

$$\left.\begin{aligned} \kappa\rho^* &= 3f^{*\prime 2}/f^{*2}, \\ \kappa p^* &= -2f^{*\prime\prime}/f^* - f^{*\prime 2}/f^{*2}, \end{aligned}\right\} \tag{7}$$

where κ is the constant of gravitation and the primes indicate differentiations with respect to t^*. We are here taking Robertson's k equal to zero, since our t-space is flat, and we are taking the cosmical constant λ occurring in Einstein's law of gravitation to be zero, since if it were not zero it would have to be very small not to be in disagreement with observation and its reciprocal would then provide us with a very large constant number, in contradiction to our fundamental principle.

Substituting (6) into (7), we get

$$\kappa\rho^* = \tfrac{4}{3}t^{*-2} = \tfrac{16}{3}t^{-4}, \tag{8}$$

$$\kappa p^* = 0. \tag{9}$$

From (2) and (3) we have

$$\rho \propto t^{-1},$$

which is in agreement with (8) when one remembers the different units of distance used in the measurement of ρ and ρ^*. This agreement should not be regarded as a support for our present theory, however, since it is due simply to the observed average density of matter being of the same order of magnitude as that to be expected from the curvature of space-time (assuming a radius of curvature of the order of the reciprocal of Hubble's constant), which fact provides a satisfactory feature in every theory of cosmology. On the other hand the result (9) may be regarded as a support for our theory, since the average hydrostatic pressure in space, due mainly to radiation pressure, is extremely small compared with the average density of matter, and so should be counted as zero in a first approximation.

. Summary

It is proposed that all the very large dimensionless numbers which can be constructed from the important natural constants of cosmology and atomic theory are connected by simple mathematical relations involving coefficients of the order of magnitude unity. The main consequences of this assumption are investigated and it is found that a satisfactory theory of cosmology can be built up from it.

References

Milne, E. A. 1936 *Proc. Roy. Soc.* A, **158**, 324–48.
— 1937*a* *Proc. Roy. Soc.* A, **159**, 171–91.
— 1937*b* *Proc. Roy. Soc.* A, **159**, 526–47.
Reynolds, J. H. 1937 *Nature, Lond.*, **140**, 387.
Robertson, H. P. 1933 *Rev. Mod. Phys.* **5**, 62–90.
Walker, A. G. 1936 *Proc. Lond. Math. Soc.* **42**, 90–127.