A radical departure from the 'steady-state' concept in cosmology*

By F. Hoyle, F.R.S. and J. V. Narlikar
University of Cambridge and the California Institute of Technology

(*Received* 1 *March* 1965)

The results in this paper are based on an entirely different choice of the undetermined coupling constant f which appears in the theory of creation of matter. Previously f was chosen to make the steady-state expansion rate coincident with the observed expansion rate. Now that we take a much larger value for f, the corresponding steady-state expansion rate is much greater than the observed value. We interpret this difference as showing that we live in a wide, possibly temporary, fluctuation from the steady-state situation. The expansion rate in such a fluctuation follows the Einstein–de Sitter relations.

The natural scale set by the new steady-state corresponds to the masses of clusters of galaxies, we obtain $10^{13}M_{\odot}$ instead of $10^{23}M_{\odot}$ for the 'observable universe'. It is suggested that elliptical galaxies were formed early in the development of a fluctuation. Our discussion of high energy phenomena leads to immediate explanations of the energy spectrum of cosmic rays, of the presence of e^+ in cosmic rays and of the rate of energy production associated with radio sources.

Introduction

In the preceding paper (Hoyle & Narlikar 1966, hereafter referred to as I), we explored the possibility that creation of matter in the universe takes place, not uniformly as required by the homogeneous steady-state theory, but in a discrete manner around isolated centres. We called these centres 'pockets' of creation; and the theory required the pockets to be associated with strong gravitational fields. It was shown, for example, that massive galaxies can act as pockets of creation and that it is possible for one generation of galaxies to reproduce another over times of the order $\frac{1}{3}H^{-1}$, H being the Hubble constant. It was also shown that the universe can be maintained in a state of steady expansion by the C field arising from creation in the pockets. In this way it was possible to establish a connection between galaxies and cosmology—clearly an advantage over the homogeneous theory which dismisses galaxies as 'local irregularities'.

In the present paper we consider further applications of the concept of pocket creation—especially in connexion with the production of high energy particles and the nature of radio sources. For this, it is necessary to summarize some of the results of paper I.

The field theoretic structure of the problem is the same as the one which leads to the homogeneous steady-state theory. Thus, the creation of matter is described by means of a scalar field C which modifies Einstein's equations by adding new terms to the right hand side

$$R_{ik}-\tfrac{1}{2}g_{ik}R = -8\pi G(T_{ik}+H_{ik}). \tag{1}$$

In (1), T_{ik} is the energy tensor of matter and H_{ik} the C-field tensor

$$H_{ik} = -f(C_i C_k - \tfrac{1}{2}g_{ik}C_l C^l), \quad C_i = \partial C/\partial x^i. \tag{2}$$

f is a coupling constant.

* Supported by the Office of Naval Research Contract Nonr-220 (47) and by the National Aeronautics and Space Administration Research Grant NGR-05-002-028.

The C field arises whenever particles are created or destroyed. It satisfies the source equation

$$C^k{}_{;k} = n/f, \tag{3}$$

where n = creation of particles per unit proper time per unit proper volume. The destruction of particles is counted as negative creation in (3). Detailed conservation laws require that, at the point of creation, the net momentum p_i of created particles (which may be in the form of baryons, mesons, leptons, etc.) satisfies

$$p_i = C_i. \tag{4}$$

Hence if E is the total energy available, we must have

$$E^2 = p_i p^i = C_i C^i \tag{5}$$

for creation to be possible.

In the homogeneous steady-state theory baryons (and leptons) of rest mass m_0 are created at rest at a uniform rate everywhere so that

$$C_i C^i = m_0^2. \tag{6}$$

In the homogeneous isotropic case, only the time derivative of C is non-zero. This leads to the steady-state line element

$$ds^2 = dt^2 - e^{2Ht}(dr^2 + r^2\,d\Omega^2), \quad d\Omega^2 = d\theta^2 + \sin^2\theta\,d\phi^2, \tag{7}$$

where
$$\dot{C} = m_0, \quad H^2 = \tfrac{4}{3}\pi G f m_0^2. \tag{8}$$

The coupling constant f and m_0 together determine the value of H. Creation is uniform and at a rate $3Hfm_0^2$ everywhere.

In the theory of pocket creation, creation takes place only in the neighbourhood of a massive object. This can happen in the following way. Suppose the cosmological level of $C_i C^i$ is a little less than m_0^2,

$$C_i C^i = m^2 < m_0^2. \tag{9}$$

Then, in the homogeneous theory there will be no creation at all. In the presence of a massive object, however, the gravitational field may be strong enough to raise $C_i C^i$ to the threshold m_0^2 for creation. It was shown in I that in the neighbourhood of a spherical object of mass M, radius R, the value of $C_i C^i$ can be raised to

$$C_i C^i = m^2(1 - 2GM/R)^{-1}, \tag{10}$$

where R is the Schwarzschild coordinate. It is assumed here that the C field included in (10) arises only from distant sources and that it does not modify the Schwarzschild line element

$$ds^2 = dT^2\left(1 - \frac{2GM}{R}\right) - \frac{dR^2}{1 - 2GM/R} - R^2\,d\Omega^2, \tag{11}$$

which holds in the neighbourhood of the object. Thus in a sufficiently strong gravitational field the creation threshold can be attained. If creation in the neighbourhood of the object is taken into account, (10) has to be modified. It can be shown that if the creation rate in the interior to R is $4\pi fmA$, then

$$\frac{\partial C}{\partial T} = m, \quad \frac{\partial C}{\partial R} = -\frac{Am}{R^2}\left(1 - \frac{2GM}{R}\right)^{-1}, \quad C_k C^k = m^2\,\frac{1 - A^2/R^4}{1 - 2GM/R}. \tag{12}$$

When (12) is used, there is a critical value of $R = R_c$ up to which creation can take place. This is given by

$$R_c = \tfrac{3}{2}GM(1-m^2/m_0^2)^{-1}. \tag{13}$$

The corresponding creation rate is

$$Q = 3\sqrt{3}\,\pi f m^2 (GM)^2 (1-m^2/m_0^2)^{-\frac{3}{2}}. \tag{14}$$

Thus, when m is close to m_0, R_c and Q can be very large. The creation rate given by (14) is the maximum possible creation rate; the actual creation rate in a particular case lies between zero and this value.

In I we were mainly concerned with galaxies and clusters of galaxies, for which the parameter GM/R_b is $\ll 1$. Accordingly, we required m to be close to m_0, for the creation rate to be appreciable. As seen from (13) and (14), both R_c and Q are very sensitive to the difference $1-m/m_0$ when it is small. In the next section we shall consider the opposite case where GM/R_b is close to unity. Such conditions are favourable for the production of high energy particles.

Creation of energetic particles

From (5) we see that, for the production of high energy particles, $C_i C^i$ must be large. This is possible, as shown by (12), provided $2GM/R$ is close to unity. The question arises, therefore, as to whether the particles created in such a strong gravitational field will have enough energy to escape to infinity. We first consider this problem.

Using the metric given by (11), the C field given by (12), we see from (4) that the momentum p^i of the baryon–lepton pair with total rest-mass m_0 created at $R = R_1$ has the following non-vanishing components:

$$p^4 = m\left(1-\frac{2GM}{R_1}\right)^{-1}, \quad p^1 = \frac{Am}{R_1^2}. \tag{15}$$

Considering the pair to follow an outward radial geodesic, with (15) as the initial conditions, we get

$$\frac{\mathrm{d}T}{\mathrm{d}s} = \frac{m/m_0}{1-2GM/R}, \quad \left(\frac{\mathrm{d}R}{\mathrm{d}s}\right)^2 = \left(\frac{m}{m_0}\right)^2 - 1 + \frac{2GM}{R}. \tag{16}$$

Since we consider $m < m_0$, it is clear that the pair cannot escape to infinity, a result anticipated in I. In fact, the radial motion becomes zero at a distance

$$R = 2GM(1-m^2/m_0^2)^{-1} = \tfrac{4}{3}R_c. \tag{17}$$

Thus, provided we are dealing with the simple metric (11), all newly created material stays within a radius slightly greater than R_c. Accumulation of matter would, however, modify M and hence the metric (11). In such a case an unstable situation develops and (11) no longer applies.

We now proceed to investigate the case of strong gravitational fields, i.e. the case where the radius R_b of the central object only slightly exceeds $2GM$. We shall assume that there is no creation inside the body, so that $A = 0$ on the surface. Consider an electron-proton pair created at $R = R_1$ near the surface. The total energy of the pair is given by

$$E^2 = C_k C^k = m^2(1-2GM/R_1)^{-1}. \tag{18}$$

Let m_p, m_e denote the masses of the proton and the electron respectively, so that $m_0 = m_p + m_e$. The sum of the kinetic energies of the two particles is $E - m_0$. For $E - m_0 \ll m_0$ the electron takes essentially the whole of the energy, whereas in the opposite case, $E - m_0 \gg m_0$, each particle has kinetic energy $\sim \frac{1}{2}(E - m_0)$. Under what circumstances can the electron escape from the object? The most favourable condition is that the electron is emitted radially outwards. Write W for the electron energy. Initially

$$m_e \, \mathrm{d}T/\mathrm{d}s = W \quad \text{at} \quad R = R_1. \tag{19}$$

If the electron moves along a radial geodesic

$$\frac{\mathrm{d}T}{\mathrm{d}s} = \frac{W}{m_e} \frac{1-2GM/R_1}{1-2GM/R}. \tag{20}$$

Inserting this in the line-element (11) (which is a first integral of the geodesic equations),

$$1 = \left(1 - \frac{2GM}{R}\right)\left(\frac{\mathrm{d}T}{\mathrm{d}s}\right)^2 - \frac{1}{1-2GM/R}\left(\frac{\mathrm{d}R}{\mathrm{d}s}\right)^2, \tag{21}$$

the condition that the electron reaches infinity is just

$$\frac{W}{m_e}\left(1 - \frac{2GM}{R_1}\right) > 1. \tag{22}$$

In the highly relativistic case

$$W \simeq \tfrac{1}{2}(E - m_0) \simeq \tfrac{1}{2}E = \frac{1}{2}\frac{m}{(1-2GM/R_1)^{\frac{1}{2}}}. \tag{23}$$

Hence (22) is equivalent to

$$\frac{1}{2}\frac{m}{m_e}\left(1 - \frac{2GM}{R_1}\right)^{\frac{1}{2}} = \frac{1}{2}\frac{m^2}{Em_e} > 1. \tag{24}$$

For the case $m \simeq m_0$, (24) gives

$$E < \sim m_p^2/2m_e. \tag{25}$$

Thus a newly created electron escapes from the associated object for all energies up to 10^3 GeV. This is just the order of the electron energies required to explain the optical synchrotron radiation from such objects as the Crab Nebula, M 87 and M 82. A corresponding analysis for protons shows that protons do not escape from the gravitational field of the associated object.

It is possible that we have here the beginnings of an understanding of the origin of the higher energy electrons, known to be present in radio sources. Several objections can be raised. A continual net loss of electrons would soon build a positive charge excess that would destroy the geodesic motion assumed above, and which would hold back all further electrons. However, even a diffuse ionized gas, present around the object, could supply an inflow of low energy electrons that would be adequate to maintain charge neutrality near the object. A more serious objection is that the emerging electrons do not reach infinity with energy W, but with energy

$W(1-2GM/R_1)^{\frac{1}{2}} \simeq \frac{1}{2}m_p$. Hence the electrons arrive at infinity with energy of $\sim 1\,\text{GeV}$, not $10^3\,\text{GeV}$. While this is of the order required to explain radio synchrotron emission from the sources, it is not sufficient to explain the optical synchrotron emission. This difficulty can be overcome if the object in question is in a state of oscillation. The parameter $2GM/R_b$ then varies through the oscillations. When it is close to unity, the pair created would have high energy. During the expansion phase of the object, the electrons are pushed out. The electrons therefore do not have to supply the whole of the energy necessary to take them out of the gravitational field of the object. In such a case energy is removed from the object whose oscillations are therefore damped. So long as $2GM/R_b$ is appreciably less than unity at the maximum radius, the electrons can escape without any great loss of their initial energy. Indeed protons can also escape with essentially equal facility. The energy extracted from the object in this way is of the order of the rest mass energy of the object, which is $\sim 10^{54} M/M_{\odot}$, adequate for supernovae when $M \sim M_{\odot}$ and for radio sources and quasi-stellar objects when $M \sim 10^7 M_{\odot}$.

Assuming the oscillation case, what is the energy spectrum of the emitted electrons and protons? To answer this question it is necessary to know whether or not the creation rate is dependent on E. For the creation of a single pair there is no factor from the momentum space, so the energy-dependence need not be strong. In the absence of a more definitive theory, we assume no dependence in the following. (Any assigned dependence could easily be inserted.) The rate of creation in a shell between R and $R+\mathrm{d}R$ is proportional to $4\pi R^2\,\mathrm{d}R$, since in the Schwarzschild line-element (11), $\sqrt{(-g)} = R^2 \sin\theta$, as in flat space. The value of E associated with R is given by $E^2 = m^2(1-2GM/R)^{-1}$ which differentiates to give

$$E\,\mathrm{d}E = -m^2\left(1-\frac{2GM}{R}\right)^{-2}\frac{GM}{R^2}\,\mathrm{d}R. \tag{26}$$

Writing $\mathrm{d}N$ for the creation rate associated with the shell between R and $R+\mathrm{d}R$, we therefore have

$$\mathrm{d}N \propto 4\pi R^2\,\mathrm{d}R \propto R^4(1-2GM/R)^2\,E\,\mathrm{d}E. \tag{27}$$

For R close to $2GM$, the R^4 factor in (27) is essentially constant as E varies. Hence the energy spectrum is given by

$$\mathrm{d}N \propto \left(1-\frac{2GM}{R}\right)^2 E\,\mathrm{d}E \propto \frac{\mathrm{d}E}{E^3}. \tag{28}$$

While this agreement with the spectrum of cosmic rays, and with that of electrons in radio sources, may be a pure coïncidence, we feel impressed by the simplicity and universality of the argument. There is no appeal to special effects and conditions, no special choice of parameters, only an assessment of the spatial volume associated with the energy range between E and $E+\mathrm{d}E$. This is proportional to E^{-3}.

In the above we have explicitly considered electrons and protons. The same arguments could equally well hold for other baryon–lepton pairs such as (Σ^+, e^-), (Σ^-, e^+), etc. Only a quantum theory of the creation process would be able to give a precise answer as to what proportion of each pair would be present in any general mixture. The possibility of meson production will also have to be taken into account.

The important point that emerges, however, is that both electrons and positrons would be present at sufficiently high energies. Since charge is conserved in the creation process, a positively charged baryon will be accompanied by an electron, and a negatively charged baryon will be accompanied by a positron. At kinetic energies below 1 GeV, higher mass pairs like (Σ^+, e^-), (Σ^-, e^+) will be less frequent and (p, e^-) will predominate. We would thus expect the proportion of electrons to be more than that of positrons and that the latter would increase at energies above ~ 1 GeV. This result agrees qualitatively with the measurements of De Shong, Hildebrand & Meyer (1964).

Once again we find the simplicity impressive. Indeed, while suggestions leading to the energy spectrum have previously been made along quite different lines from those above, no other explanation has yet been made for the observations of De Shong *et al.*, except one which involves an artificial coincidence. It was pointed out some years ago by Ginzburg that it would be possible to distinguish between the two possibilities: (*a*) that high energy electrons of the radio sources are primary, in which case it was thought that the electrons would be entirely e^-; (*b*) that the electrons are secondaries from nuclear collisions of cosmic-ray protons, in which case there should be an excess of e^+. The observation that positrons are present but not in excess fits neither (*a*) nor (*b*). A strange situation is required in which both primary and secondary electrons are needed and the two processes (*a*) and (*b*) are closely comparable. So far as we are aware, no explanation has been given for this *ad hoc* coincidence. The above derivation seems to us much more satisfactory.

The chemical composition of cosmic rays is quite unlike that of any material with which we are acquainted in astrophysics. The excess heavy element concentrations, as compared to ordinary stellar material, have been held to point to supernovae as the sources of cosmic rays, because heavy elements are probably synthesized inside the particular stars that become supernovae. However, samples of material in which heavy elements have been synthesized do not contain hydrogen. Mixing of heavy elements from the interior to the surface, followed by acceleration of surface material, seems necessary in the supernova picture. So in this picture the chemical composition of cosmic rays is determined by the more or less accidental features of the mixing process, and the fact that the total energy of cosmic ray protons is approximately the same as the total energy of all other nuclei becomes another coincidence.

In our picture there is nothing special about protons. At high values of E, more complex particles, decaying to heavy nuclei, can be created—in a sense we are then dealing with multiple pair production. A more detailed theory is evidently needed to work out relative creation rates. However, on qualitative grounds we expect:

(i) the energy spectrum to be the same for all particles since the spectrum arises from essentially geometrical considerations, without reference to the kind of particle concerned;

(ii) under conditions of energy excess there will be discrimination against weakly bound nuclei, D, He^3, Li, Be, B; and strongly bound nuclei, particularly He^4, will be favoured;

(iii) a relation between the total proton energy and that of other nuclei.

It is of interest that the theory predicts the presence of D, ^{3}He in the primary cosmic rays. We would expect

$$\frac{^3\mathrm{He}}{^4\mathrm{He}} \approx \frac{\mathrm{Li, Be, B}}{\mathrm{C, N, O}}, \tag{29}$$

lightly bound nuclei being in the numerators in both ratios and comparatively strongly bound nuclei in the denominators. Although we are not in a position to discuss (iii) on a quantitative basis, it is no surprise that the two energies are comparable.

Cosmology

The theory described in the previous sections is plainly an improvement over the homogeneous steady-state theory. It not only connects galaxies and cosmology, as shown in the previous paper; but it also explains the origin of high energy particles. However, one quantitative aspect remains to be investigated, namely the creation rate of high energy particles. This should match the rate of energy output of radio sources and of quasistellar objects in order that the theory be satisfactory.

Although the concept of maximum creation rate gives an upper limit to the creation rate in a strong gravitational field, the rate could certainly be of this general order. To estimate it we proceed as follows. From (12), we see that, in order to keep $C^iC_i > 0$, we must have $A < R^2$ for $R > 2GM$. Since high energy particles are created near $R = 2GM$, A cannot exceed a quantity of the order $4G^2M^2$, giving a creation rate $16\pi fmG^2M^2$ particles per unit time. Since the energy spectrum $\mathrm{d}E/E^3$ contains most energy at the least value of E, which for energetic particles is $E \simeq m_p$,

$$\text{the rate of energy production} \sim 16\pi f m_p^2 G^2M^2 \tag{30}$$

in which we have put $m = m_p$. Next, $fm_p^2 \simeq 3H^2/4\pi G$, from (8) where $m_0 \simeq m_p$.

$$\text{So the rate in question} \sim 12H^2GM^2 \text{ per unit time.} \tag{31}$$

This is the rate in time units such that $c = 1$. Equation (31) can be written in the numerical form

$$\text{rate} \sim 10^{15}(M/M_\odot)^2 \,\mathrm{erg\,s^{-1}}. \tag{32}$$

For the Crab Nebula the energy output from synchrotron radiation is at least $10^{36}\,\mathrm{erg\,s^{-1}}$. Since the mass in this case cannot be very large compared to $M_\odot$, it follows that (32) is too small by a factor of $\sim 10^{20}$. For a radio galaxy M probably lies in the range 10^6 to $10^8 M_\odot$, with perhaps a preference for the lower value (Fowler & Hoyle, in press). Then (32) gives 10^{27} to $10^{31}\,\mathrm{erg\,s^{-1}}$, again small by a factor of $\sim 10^{20}$. Manifestly the theory collapses in ruins.

Short of abandoning the whole theory, it is necessary either

(i) to drop the idea that the origin of cosmic rays and high energy electrons is connected with the creation process, or

(ii) to increase the coupling constant f by a very large factor, of the order 10^{20}.

In the first case the attractive features discussed in the preceding section are lost. The theory remains much the same as it was before—that is to say, the cosmological aspects of the theory survive as set out in I. Creation of matter is then confined to galaxies and clusters of galaxies and is at a gentle rate.

If the second possibility is considered, f will have to be increased. This means $\dot{C}$ must be reduced by $\sim 10^{-10}$ in order to maintain $f\dot{C}^2$ at the same cosmological value as before. Thus m is reduced by $\sim 10^{-10}$ and the requirement $m_0 \simeq m$ can no longer be maintained. In other words the work described in I is lost.

The situation is that we cannot retain the results of the previous section and of paper I. A choice must be made. The traditional viewpoint of the steady-state theory suggests that we take up the first possibility—that discussed in I. But prejudice apart, the empirical facts suggest the opposite. The second possibility gives agreement with actual data concerning cosmic rays and radio sources, whereas we have no direct evidence of the gentle kind of creation implied by the first possibility. Clearly then, we must follow the second possibility, even though this means throwing overboard the usual framework of the steady-state theory.

There is no requirement, however, that the universe originated in a singularity, as in the classical Friedmann cosmologies. A 'steady-state' situation is possible, with H given by $fm_p^2 = 3H^2/4\pi G$, corresponding to a density $\sim 10^{20}$ times greater than the present density—i.e. $\sim 3\times 10^{-9}\,\mathrm{g/cm^3}$ instead of $\sim 3\times 10^{-29}\,\mathrm{g/cm^3}$. Why was this steady-state abandoned? Departure from the steady-state is not possible in the homogeneous theory, but when there are inhomogeneities the average creation rate can deviate from the steady-state value. As shown in I the mass M in a pocket grows at a rate proportional to M^2. If the overall rate rises above the steady-state value, the universe simply expands faster and tends to reduce the creation rate, thus setting up an osculating steady state. But should the creation rate fall (for example, by the inhomogeneities dividing into fragments so that the M^2 dependence suddenly reduces the creation rate) an instability can develop. A drop in the creation rate reduces the overall level of C^iC_i in the universe, i.e. m is reduced. Since the creation rate is very sensitive to the difference (m_0-m), most of the pockets go out of action; only those with a strong gravitational field can still produce particles. This means that no new pockets are formed, while the old ones are expanded away. No osculating steady-state is therefore set up under such circumstances. Given sufficient time we expect this situation to arise, *although not everywhere throughout the universe synchronously,* since the pattern of inhomogeneities will not be the same everywhere. We therefore need consider only finite portions of the universe where the creation rate is reduced effectively to zero.

The behaviour of such finite portions can be described by means of the Robertson–Walker line element

$$ds^2 = dt^2 - S^2(t)\left[\frac{dr^2}{1-kr^2} + r^2\, d\Omega^2\right], \tag{33}$$

in which the coordinate r is the same intrinsic radial coordinate as was used in the previous steady-state situation and k is a small positive constant whose value is determined by the size and nature of the instability. In the absence of creation we have

$$\dot{C} = \frac{A}{S^3}, \quad \rho = \frac{B}{S^3} \quad (A, B \text{ constants}), \tag{34}$$

and

$$S^4\dot{S}^2 = \tfrac{8}{3}\pi G(BS^3 - \tfrac{1}{2}fA^2) - kS^4, \tag{35}$$

with B, f, k all positive. For large S the k term dominates the expansion and reduces $\dot{S}$ to zero. The finite portion in question therefore expands like a bubble but then falls back on to the 'steady state'. Bubbles may occur at any place and time, but need not develop synchronously. Bubbles develop, not because of a more rapid expansion than in surrounding regions, but because the creation process is cut off inside them. The C field propagates from the surrounding regions into the bubble and increases $C_i C^i$, tending to re-establish creation. This process, however, involves a surface to volume effect. Small bubbles will be 'filled in' quickly and large ones more slowly. Hence we can set a lifetime to any finite instability that may develop.

The condition that a bubble has not yet filled in is that a signal, travelling on a null geodesic, emitted at the boundary at the moment the bubble began to evacuate, must not yet have had time to reach the central regions. To investigate this problem take $S = 1$, $t = 0$, at the moment evacuation began. Then the constant B is the density at $t = 0$, i.e. the steady-state density, $\sim fm_p^2$. As the bubble expands the A^2 term in (35) becomes negligible, and so long as the k-term is small, we have

$$S\dot{S}^2 \simeq \tfrac{8}{3}\pi G f m_p^2, \tag{36}$$

which integrates to give

$$S = (6\pi G f m_p^2)^{\frac{1}{3}} t^{\frac{2}{3}}. \tag{37}$$

To avoid confusion about the meaning of H, we define the instantaneous value of $\dot{S}$ by $\mathscr{H}(t)$, so that from (37),

$$\frac{\dot{S}}{S} = \mathscr{H}(t) = \frac{2}{3t}. \tag{38}$$

H continues to have the old meaning, given by $fm_p^2 = 3H^2/4\pi G$. Equation (38) is the usual Einstein–de Sitter result.

A simple calculation then shows that a signal emitted at $t = 0$ from a particle with radial coordinate r reaches the observer at $r = 0$ at a time t given by

$$rS(t) = 2\mathscr{H}^{-1}. \tag{39}$$

$rS(t)$ represents the distance of the emitting particle from the central observer at time t. Apart from a factor 2, this is the same as the distance of the event horizon in the old steady-state theory and is therefore of order 10^{28} cm. The bubble we live in must be at least as large as this and must contain a mass of the order of $10^{23} M_{\odot}$, otherwise the bubble must have filled before now.

For any particular bubble there is a maximum r, that associated with the boundary. Hence it will eventually be filled in. This is the 'surface effect' described before and is associated with the k term. The k term involves one more power of S than the density term in (35) and therefore represents a surface effect. The above physical argument is given to show that the magnitude of the k-term depends on the size of the bubble and vice versa.

Our picture then is of a 'steady-state' universe with average density $\sim 10^{-8}$ g/cm^3, some 10^{20} times higher than the average density in the old steady-state theory. Inhomogeneities play an important role in the manner described before. Inhomogeneities can lead to instabilities developing, the instabilities being regions that become evacuated because the creation process is temporarily cut off. Such bubbles eventually fill in, the filling in process being quicker for smaller bubbles than for the larger ones. We are living in such a bubble which has not yet filled in.

The argument we gave in a previous paper (Hoyle & Narlikar 1963) concerning the asymmetry of time, the consistency of retarded solutions of Maxwell's equations, but not of advanced solutions, survive essentially unchanged in this picture, since the universe as a whole is in a steady state. Our argument could fail, however, if there was any possibility of the whole universe getting out of hand. What would happen if in some way instability managed to develop synchronously everywhere?

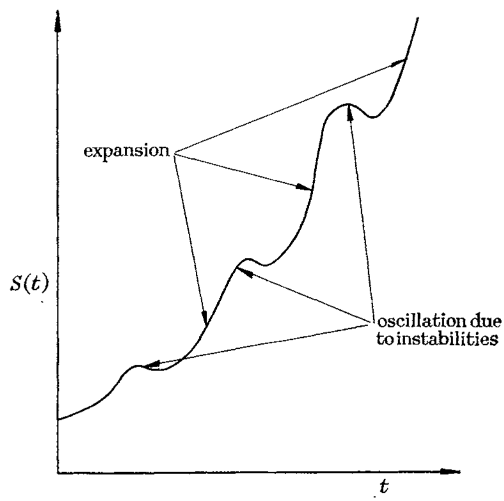

FIGURE 1

Would the universe then expand to zero density in the fashion of the Einstein–de Sitter cosmology? The answer must be affirmative unless the equation for $\dot{S}$ develops a zero for large S. In the case of a finite portion of the universe this is possible by means of a k term, as seen above; for the whole universe, however $k = 0$. If we alter the topology of the universe by setting $k = 1$, this would make the universe fall back into a contracting phase as in the case of a bubble. Such a contraction would continue, until the universe is made to bounce by the growing $\dot{C}^2$ term. During its re-expansion, the value of $\dot{C}^2$ drops to the threshold level, and the creation process reassert itself, giving a return to the steady-state situation. This would continue until the next instability arose. The universe would follow the kind of expansion-contraction phase shown in figure 1, with the expansion dominating.

To complete this aspect of our discussion, we note that, even in the case $k = 0$, it may still be possible to prevent the expansion of the universe to zero density, if we take into account the result that creation still goes on in a strong local gravitational field. The mass M in such a pocket creates particles at a rate proportional to M^2. Such an instability is capable of raising the creation rate to very high values, if the mass M does not break up. In such a case, the $\dot{C}^2$ value will be raised. Once it is raised to the threshold required for creation in weak gravitational fields, the universe would again attain a steady state. However, such a process requires an enormous increase in the production of high energy particles, and, as a result, a fantastic increase (by a factor of $\sim 10^{24}$) in the energy density of cosmic rays above the present value.

Galaxy formation

In order to introduce a very attractive feature of the present theory we repeat an interesting argument from the old theory. It is always possible to take $G = 1$ as well as $c = 1$. Then, instead of the usual mass, length and time units, we only have a length unit. In order to define the number of particles in a galaxy, it is necessary to obtain a large dimensionless number in some way, a number of the order 10^{70}. In Friedmann cosmologies the only dimensionless numbers available are those obtained by comparing masses of different particles—and these do not give 10^{70}. When creation is introduced, the coupling constant f appears with dimension $(\text{length})^{-4}$, and a new ratio is therefore obtained. Choosing the unit of length so that $m_p = 1$, we have $f = 3H^2/4\pi \simeq 10^{-160}$. The length unit associated with f was therefore taken as 10^{40}, i.e.

$$f^{-\frac{1}{4}}/m_p \simeq 10^{40}. \tag{40}$$

A large dimensionless number appears, but it does not lead to 10^{70} in a simple way. Its square, $\sim 10^{80}$, the value of H^{-1}, is taken to represent the number of particles in the observable universe. The length contained in f was of cosmological significance and not applicable to galaxies.

The length contained in f was reduced by way of a 'hot universe' (Hoyle 1958) in the following way. Newly created baryons were taken to be neutrons, which decayed, releasing an average kinetic energy $\sim \frac{1}{5}m_e$ per particle. Pressure fluctuations were capable of combating expansion over distances of the order $(m_e/5m_p)^{\frac{1}{2}} H^{-1}$. This was taken as defining the initial linear scale of the condensations that went to form superclusters of galaxies. The masses depended on the volumes of the condensations and hence on $(m_e/5m_p)^{\frac{3}{2}} \simeq 10^{-6}$. The corresponding factor when newly created baryons were taken to be an equal mixture of neutrons and protons was $\sim 10^{-7}$. The argument fails because the X-ray background associated with hot intergalactic material has not been observed. Failure came as something of a disappointment because the argument is in principle a reasonable one.

Returning to the cosmology described in the previous section, we see that f is now increased by $\sim 10^{20}$ over its value in the old steady-state theory. With this new value of f, we have

$$f^{-\frac{1}{4}}/m_p \simeq 10^{35}, \tag{41}$$

and the corresponding H^{-1} is $\sim 10^{70}$.

These considerations can be restated in more familiar language as follows. With the new value of f, the radius of the observable universe and its mass are both reduced by a factor of $\sim 10^{10}$, the latter to a value $\sim 10^{13}M_{\odot}$. This mass defines a typical condensation: a massive elliptical and a score of spirals. H^{-1} is the natural communication length at the beginning of the development of the bubble. As the bubble expands, the communication length increases as $\mathscr{H}^{-1}$. The present-day length $\mathscr{H}^{-1}$ has naturally nothing to do with the length associated with galaxies.

A dimensionless number of the same order as (41) arises when we compare the gravitational and electromagnetic interaction between two protons

$$e^2/Gm_p^2 \simeq 10^{36}. \tag{42}$$

We could therefore attempt to argue that the formation of galaxies must be a process that relates electromagnetic forces to gravitation. Electromagnetic forces appear as pressure gradients, as in the hot universe, with the difference, however, that the observational difficulty concerning the X-ray background intensity can perhaps be avoided by placing the epoch of condensation in the remote past. What criterion determines this epoch?

Many years ago, Gamow pointed out that there is a unique moment in a Friedmann cosmology at which the energy density of radiation can be equal to the energy density of matter. So far as we are aware, this is the only suggestion for defining an epoch of condensation. Let ρ_0, T_0 denote the present-day mass density and temperature of the radiation field. Then in the past

$$\frac{T}{T_0} = \left(\frac{\rho}{\rho_0}\right)^{\frac{1}{3}}. \tag{43}$$

For the radiation density and mass density to be equal,

$$aT^4 = \rho, \tag{44}$$

in which a is the radiation constant. For ρ in g/cm^3 and T in °K, (44) takes the numerical form

$$T \simeq 10^9 \rho^{\frac{1}{4}}. \tag{45}$$

Eliminating T between (43) and (45) gives

$$\rho = (10^9 \rho_0^{\frac{1}{3}}/T_0)^{12}, \tag{46}$$

a result very sensitive to T_0. For $\rho_0 \simeq 10^{-29}\,\text{g/cm}^3$, $\rho \simeq 10^{-2}\,T_0^{-12}\,\text{g/cm}^3$. Attempts have usually been made to make ρ come out at a typical galactic density, of the order $10^{-24}\,\text{g/cm}^3$. This requires $T_0 \simeq 10^2\,°\text{K}$, a value too high according to the radio data, which sets the present-day radiation temperature not above 10 °K. Most radio astronomers favour a value close to 1 °K. The discrepancy between theory and observation arises because it is usually assumed that the linear scale of galaxies is very small compared to $\mathscr{H}^{-1}$ at the epoch of condensation. As our theory shows, $\mathscr{H}^{-1}$ should be the length scale to be associated with the galaxies. At the time of condensation $\mathscr{H}^{-1} = H^{-1} \simeq 4 \times 10^7\,\text{s}$, and $\rho \simeq 10^{-8} g/\text{cm}^3$. From (46) we get $T_0 \simeq 1\,°\text{K}$, consistent with the radio observations.

A theory similar to this could be worked out for a Friedmann cosmology, if a satisfactory explanation is given of why the moment of equality of radiation and matter energy densities should be of critical importance. For a density $\rho \simeq 10^{-8}\,\text{g/cm}^3$ the mean free path of radiation is very small compared to $\mathscr{H}^{-1}$, in fact $\rho\mathscr{H}^{-1} \simeq 10^{10}\,\text{g}$. Hence we cannot expect fluctuations in the radiation field over a scale $\mathscr{H}^{-1}$ at this epoch. Suitable fluctuations in mass distribution are therefore the only way to provide any significance to this epoch. This can be done by assuming considerable initial inhomogeneities in the matter distribution, but a homogeneous radiation field. Initially the latter dominates and the geometry is homogeneous. At, say, $\rho \simeq 10^{-8}\text{g/cm}^{-3}$, the matter terms become more important; at this epoch the inhomogeneities tend to separate out. In this way it may be possible to obtain a theory of galaxy formation within the framework of a Friedmann cosmology. Such a theory would not, however, explain the high energy phenomena as the theory presented here does.

The theory solves perhaps the most puzzling property of galaxies, the existence of what seems to be a strict upper limit to their masses and luminosities. This is the limit set by H^{-1}, with H given by $3H^2/4\pi G = fm_p^2$. The distance H^{-1} determines the communication length at the onset of evacuation of a bubble. However, more precisely, what are the factors determining whether or not a particular sample of material goes to form a galaxy? The present-day average density of condensed matter cannot be greater than $\sim 10^{-30}\,\mathrm{g/cm^3}$, whereas the theory requires the total present-day intergalactic matter density to be $\sim 10^{-29}\mathrm{g/cm^3}$. Hence we must conclude that some special condition had to be satisfied in order that a galaxy be formed from a particular sample of material. It is natural to appeal to inhomogeneities already present at the beginning of evacuation. An interesting situation arises if we postulate that the excess of mass takes the form of a condensed object of mass $\mu \ll \rho H^{-3}$, the gas being otherwise of uniform density fm_p^2. In the absence of such a central mass, the gas in the neighbourhood of the centre has just enough energy to expand away to low density as the bubble proceeds to evacuate itself according to the Einstein–de Sitter law. The effect of the mass μ is to eventually pull back the expanding gas. This will be the case for all gas that is sufficiently close for the problem to be considered a Newtonian one. For gas that is sufficiently far away, non-Euclidean terms will be more important. For this gas the mass μ does not have sufficient restraining power, and it is therefore lost. The scale factor over which non-Euclidean terms become important is H^{-1}. The situation is that as μ increases, its restraining power increases; but not unless $\mu = \rho H^{-3}$ can it restrain gas over the distance H^{-1}. Numerically, we have

$$\rho \simeq 10^{-8}\,\mathrm{g/cm^3}, \quad H^{-1} \simeq 10^{18}\mathrm{cm}, \quad \rho H^{-3} \simeq 10^{13} M_\odot. \tag{47}$$

Evidently, $10^{12} M_\odot$ is a reasonable upper limit for the mass, and to obtain a value as this μ might have to be $\sim 10^9 M_\odot$. This and other questions relating to the formation of galaxies will be discussed in a subsequent paper.

Dimensionless ratios of coupling constants

We add a brief comment on the coupling constants to be found in different parts of physics. We have already noted the coincidence between

$$e^2/Gm_p^2 \simeq 10^{36}, \tag{48}$$

$$f^{-\frac{1}{2}}/m_p \simeq 10^{35}, \tag{49}$$

which can be expressed as

$$e^2 f^{\frac{1}{2}}/Gm_p \simeq 10. \tag{50}$$

To this we can add two further ratios constructed from the β-decay constant $\mathscr{G}$ and the atomic constant $\hbar$

$$\hbar/e^2 = 137{\cdot}039, \tag{51}$$

$$e^2/m_p^2\,\mathscr{G} = 10^3. \tag{52}$$

Depending on precise formulation, a strong coupling constant could be introduced and a dimensionless number analogous to the above, and not greater than 10^3, would be obtained. The dimensionless numbers of physics appear therefore to span the range 1 to 10^3 (or 10^{-3} to 1 if reciprocals are taken). Is this an accident or is there some, as yet unknown, connexion between the different interactions of physics?

The numerical coincidence (50) (which is more striking with the new value of f than with the old) points strongly to a direct connexion between the C field and the electromagnetic field, suggesting that e and f are not independent constants. In a previous paper (Hoyle & Narlikar 1964*a*) we pointed out that the three action terms for mass (gravitation), electromagnetism and the C field have strong formal similarities and we suggested that a unified theory of gravitation and electricity might be obtained by collapsing the first two terms of the action into a single term. What the coincidence of (48) and (49) suggests is that the C-field action should also be joined with the other two in a single term. We shall attempt to develop this point of view in a future paper.

High energy phenomena

The above considerations suggest that the massive objects in radio sources and quasars are condensed residues from the true steady-state situation. Their scale is considerably less than $H^{-1} \simeq 10^{18}$ cm and their masses could range up to perhaps $10^9 M_\odot$. This statement concerning size refers to massive interiors, not to external clouds that may form about them, and which in present-day circumstances could be ten or more parsecs in diameter.

The relativistic parameter $2GM/R$ becomes of order unity at a density

$$\sim 2\times 10^{16}(M_\odot/M)^2\,\mathrm{g/cm^3}. \tag{53}$$

Clearly no great measure of contraction, from an initial density of $\sim 10^{-8}\,\mathrm{g/cm^3}$, is necessary to attain the density given by (53) when M is in the range 10^5 to $10^9 M_\odot$. There seems no reason why rotation should impede the collapse of a condensation towards a situation in which $2GM/R$ becomes of order unity.

Implosion into a singularity is prevented by the C field, in accordance with an earlier discussion. Inside the object we must have

$$f\dot{C}^2 \simeq 10^{16}(M_\odot/M)^2\,\mathrm{g/cm^3}, \tag{54}$$

whereas outside the object, in the steady-state situation, we have

$$f\dot{C}^2 \simeq 10^{-8}\,\mathrm{g/cm^3}, \quad \dot{C}^2 = m_p^2. \tag{55}$$

The second condition in (55) follows because the threshold condition $\dot{C}^2 = m_0^2$ is satisfied in the steady state, and $m_p \simeq m_0$. The first condition in (55) depends of course on our choice for the value of f. Combining these conditions, $f \simeq 10^{-8} m_p^{-2}$, and substituting in (54) gives

$$\dot{C}^2 \simeq 10^{24} m_p^2 (M_\odot/M)^2, \tag{56}$$

so that $\dot{C}^2$ is greater inside the object than it is outside by the factor $10^{24}(M_\odot/M)^2$.

These statements are based on a complete solution obtained in a former paper (Hoyle & Narlikar 1964*b*) for the case of a static body supported against implosion by the C field, with a steady-state situation at distance $H^{-1} \gg R_b$ from the body. The meaning of $\dot{C}$ is the derivative of C with respect to proper time, in (54) proper time for an observer on the body, in (55) proper time for an observer at appreciable distance form the body. Time on the body and time outside differ by the scale factor $(1-2GM/R_b)^{\frac{1}{2}}$, and this is also the factor by which the internal and external values of $\dot{C}$ differ. That is to say we require

$$\frac{1}{1-2GM/R_b} \simeq 10^{24}\left(\frac{M_\odot}{M}\right)^2 \tag{57}$$

for consistency with (56). This indeed is the relation given in equation (82) of our previous paper.

In an earlier section we found that particles of energy $m_p(1-2GM/R_b)^{-\frac{1}{2}}$ could be created near a massive body, but in our former discussion we assumed that the factor $(1-2GM/R_b)^{-\frac{1}{2}}$ could be arbitrarily large. Now we see this is not the case. Associated with a body of given M there is an upper limit of energy of approximately

$$10^{12} m_p \frac{M_\odot}{M} \simeq 10^{21}\frac{M_\odot}{M}\,\text{eV}. \tag{58}$$

For M ranging from $10^5 M_\odot$ to $10^7 M_\odot$ the upper limit ranges from 10^{16} eV down to 10^{14} eV. It is of interest that the energy spectrum of cosmic rays is believed to steepen in just this energy range. According to (58) the highest energies, $\sim 10^{20}$ eV, must be obtained from masses of stellar order. In this paper so far we have said nothing about radio sources associated with supernovae. If, at the end of stellar evolution, bodies for which $2GM \sim R$ are formed in some cases, as was proposed by Hoyle, Fowler, Burbidge & Burbidge (1964), exactly the same considerations may be applied. We could have $\sim 10^8$ stellar remnants, with masses up to $10M_\odot$, per galaxy. Because the creation rate depends on M^2 for each body, the total creation rate associated with 10^8 bodies of mass $10M_\odot$ is the same as one body of mass $10^5 M_\odot$. Compared to a single body of mass, say $10^7 M_\odot$, the combined effect of all supernova remnants in a galaxy would be weaker by a factor of $\sim 10^{-4}$. Their effect on the total cosmic-ray distribution would appear only at the highest energies. This situation is an ironic inversion of what has usually been supposed.

One of us (J. V. N.) is indebted to the California Institute of Technology for the award of a Research Fellowship and to Professor W. A. Fowler for the hospitality extended to him at the Kellogg Radiation Laboratory.

References

Hoyle, F. 1958 *Proc. 11th Solvay Conf.*, p. 53. Brussels: Institut International de Physique, Solvay.
Hoyle, F., Fowler, W. A., Burbidge, G. R. & Burbidge, E. M. 1964 *Astrophys. J.* **139**, 909.
Hoyle, F. & Narlikar, J. V. 1963 *Proc. Roy. Soc.* A, **277**, 1.
Hoyle, F. & Narlikar, J. V. 1964*a* *Proc. Roy. Soc.* A, **282**, 191.
Hoyle, F. & Narlikar, J. V. 1964*b* *Proc. Roy. Soc.* A, **278**, 465.
Hoyle, F. & Narlikar, J. V. 1966 *Proc. Roy. Soc.* A, **290**, 143.
De Shong, J. Jr., Hildebrand, R. H. & Meyer, P. 1964 *Phys. Rev. Lett.* **12**, 3.