SCIENCE

VOL. 76 — FRIDAY, SEPTEMBER 9, 1932 — No. 1967

SCIENCE: A Weekly Journal devoted to the Advancement of Science, edited by J. MCKEEN CATTELL and published every Friday by

THE SCIENCE PRESS

New York City: Grand Central Terminal

Lancaster, Pa. — Garrison, N. Y.

Annual Subscription, $6.00 — Single Copies, 15 Cts.

SCIENCE is the official organ of the American Association for the Advancement of Science. Information regarding membership in the Association may be secured from the office of the permanent secretary, in the Smithsonian Institution Building, Washington, D. C.

THE EXPANDING UNIVERSE[1]

By Dr. H. P. ROBERTSON

ASSOCIATE PROFESSOR OF MATHEMATICAL PHYSICS IN PRINCETON UNIVERSITY

THE considerations on which I am to address you this evening deal with questions which have long been of interest to the more inquisitive of mankind, questions to which answers must have been sought in that dim past in which man became the first animal capable of extended thought. The structure and meaning of that vaster world of heavenly objects gave rise to speculations, many of which have played decisive rôles in the development of civilizations and cultures. The unaided eye of the ancients limited them essentially to conjectures concerning our immediate neighbors, the other members of the solar system, and those less immediate neighbors, principally stars and configurations of stars and nebulosities, which constitute our galactic system. Only within the few centuries characterized by modern science has the telescope enabled man to explore more thoroughly that larger universe of which our own stellar system is but a member and, together with the still more recent development of the spectroscope, enabled him to bring order into apparent chaos. But the final proof that the great nebulae which have been the subject of speculation for three centuries do in fact constitute island universes comparable with our own galaxy has only been obtained within our own age, and the proof of the regularity of their distribution in space and of their relative motions is a result of the research of the past decade. These discoveries have revived old questions in a new form, and I propose this evening to set forth the partial answers which are offered by relativistic cosmology, that offshoot of the general theory of relativity which deals with the structure of the universe as a whole. But let us first briefly review the facts with which we can start and which are to be brought into order.

[1] An address delivered before the ninth annual meeting of the West Virginia Academy of Science at Athens, West Virginia, April 29, 1932.

Astronomical research has shown that our own sun is a relatively unimpressive member of that great galaxy whose lenticular shape is revealed to us, in virtue of our rather central position in it, by that striking band of stars, the Milky Way, which encircles the heavens. Our nearest neighbor, Proxima Centauri, is some four light-years removed, and on the average there is one star in every 350 cubic light-years in our neighborhood. The density of stars decreases as we go out toward the boundary of our galaxy; in any direction in the plane of the Milky Way it drops to one-hundredth of the above value at a distance of about 27,000 light-years, and in directions perpendicular to this plane the same decrease is attained at a distance of 5,000 light-years. Within these limits are a large number of non-stellar objects, such as the globular clusters and the diffuse luminous bodies known as planetary nebulae, but since we are interested here in the universe as a whole and not in the phenomena which occur within the borders of such a closely-knit system as our galaxy we pass on to the great nebulae which may be considered as counterparts of our own system.

The great extra-galactic nebulae are scattered quite uniformly over the celestial sphere, provided we attribute their apparent relative scarcity near the Milky Way to the obscuring effect of matter in our own galaxy. The distances of some of the nearer of these objects can be computed on the hypothesis that certain types of stars which are observed in them are of the same physical constitution as those which are in our own system; the resulting regularity thus found in their actual size and luminosity leads to the hypothesis that the apparent differences in size and brightness observed in this class of objects are due primarily to their distances. One of the nearest and most striking, the great spiral nebula in Andromeda, is estimated by Hubble to be some 800,000 light-years away, and its linear dimensions to be about one half those of our own galaxy. The results of surveys of extra-galactic nebulae, by Hubble at Mt. Wilson and Shapley at Harvard, indicate that although the nebulae often occur in clusters yet on a still larger scale they are fairly uniformly spaced, their average distance apart being somewhat less than 2 million light-years. The faintest of these objects, which can be well observed with the 100-inch telescope at Mt. Wilson, are of about 19th total magnitude, and are accordingly estimated to be at a distance of some 300 million light-years. Hubble estimates that about 30 million nebulae are contained within a sphere of this radius, and that so far as the observations go they are uniformly distributed, much as the molecules of gas in a container, throughout the observable universe.

In order to complete this picture of a universe consisting of nebulae fairly uniformly spaced at distances of 2 million light-years apart a knowledge of their velocities is essential. Now in discussing velocities it is to be remembered that the theory of relativity places an upper limit of 300,000 kilometers per second (the velocity of light) on the velocities of material bodies, but also that we can give no reason why the relative velocities of physically unrelated bodies should not have any value up to this limit. The relative velocities of various members of our own system may be as great as several hundred kilometers per second, but are even so but a small fraction of the limiting velocity. The measurement of velocities of objects as distant as the great nebulae is almost entirely restricted to the radial velocity, that component in the line of sight; this aspect of the motion gives rise to a Doppler effect which can be measured by the spectroscope regardless of the distance of the source, provided only that the source is sufficiently intense. The radial velocities of about 90 nebulae have been determined in this way, and it is found that all but five of these are moving away from us. Furthermore, the more distant the nebula the greater its velocity of recession; it is to be noted that the five exceptions are all our more immediate neighbors and are mainly due to the motions of the sun with respect to the galaxy. This relation between average velocity and distance is one of direct proportionality, the velocity increasing 1,000 kilometers per second for each additional six million light-years. One nebula with a velocity of recession of 19,700 km./sec. has recently been observed, corresponding to a distance of 105 million light-years. These facts indicate that there is, superposed on the relatively small peculiar motions of the nebulae, a velocity of recession which is proportional to distance and which leads to the remarkable conclusion that the universe is expanding at such a rate as to double the distance to any nebula every 1,400 million years.

This evening I wish to trace the steps which have led to an explanation of these startling observations, and I hope to show you that this explanation follows naturally from the general theory of relativity with the aid of a very few reasonable *a priori* assumptions. But let us first review those elements of the relativity theory which are essential to our argument. The Newtonian theory conceived the 4-dimensional continuum of space and time as split up into a 3-dimensional Euclidean space and a universal time which was the same for every observer, regardless of his motion. The matter immersed in such a space-time had no effect on its properties, and the motion of matter was attributed to forces, the principal one of interest at the moment being the universal force of gravitation. Out of the conflict which grew up between this picture of the world and that of the electromagnetic theory emerged Einstein's special theory of relativity. According to this theory each observer still split up the world of space and time into a Euclidean space and a "local" time, but the manner in which this division took place was dependent on the observer; each chose his reference system in such a way as to be at rest with respect to it. This transition from a geometrical to a kinematical view of space-time left essentially unaltered the status of forces, including that of gravitation. The final step, that from the special to the general theory of relativity, effected a great unification; gravitation was incorporated into the geometry of space-time. Under the general theory the special theory is still valid in any sufficiently limited portion of the space-time universe, but its geometry on a larger scale is determined wholly by the matter which it contains. In this way the brilliant program suggested by Riemann some fifty years before was carried through by Einstein.

With these facts in mind let us now return to the problem of determining the properties of a universe in which matter is, in the large, uniformly distributed in space, ignoring the irregularities in its structure due to the agglomeration of matter into nebulae. But we must first examine in what sense we may speak of a uniform spatial distribution of matter if space as such is relative to the observer in question. Does this mean that we may reinstate a universal time which is of significance to all observers? I believe that it does, within limits, and the justification of this procedure is one of the most important steps in our exposition. The only case in which failing to take account of the difference between the local reference systems of neighboring observers can lead to serious discrepancies is that in which their relative velocity is an appreciable fraction of the velocity of light. But the observations we have set forth indicate that except for relatively unimportant peculiar motions the relative velocities of neighboring nebulae are small compared with this limiting velocity, and hence observers stationed on these nebulae may choose a common reference system without undue warping. We may thus set up a reference system which splits up the universe into a cosmic space and a cosmic time which will serve in any portion of it as a mean local space and mean local time for all the observers concerned. It is to be noted that there is no contradiction between this view of a reference system of universal validity, with respect to which each nebular observer is approximately at rest, and the fact that widely separated nebulae have a considerable relative velocity; we need only refer to the very analogous situation of observers situated at points of given latitude and longitude on an expanding sphere—each of them is at rest with respect to the reference system on the sphere, but their relative distances as measured on the sphere are increasing directly with the radius.

We are now in a position to restate our facts and hypotheses in a form which leads directly to the desired solution. We have introduced a cosmic time in a way which makes significant the statement that the observed distribution of nebulae is spatially uniform, and we assume that this uniformity extends indefinitely beyond the portion of the universe to which we have observational access. (It is to be noted in passing that our procedure is to some extent equivalent to Weyl's assumption that all matter in our universe has always maintained a physical unity, even in the remote past—that the actual universe is not the fortuitous superposition of two or more incoherent parts; only with the aid of this assumption or one equivalent to it can one hope to establish a unique velocity-distance relationship.) But then according to the theory of relativity the geometry of the universe in the large must exhibit the same uniformity as the material content by which it is conditioned, and since in our idealization matter is distributed uniformly throughout cosmic space this latter must itself be homogeneous and isotropic—it must exhibit the same intrinsic properties in every point and in every direction. Now it has been known for almost a century that there are but three types of completely homogeneous and isotropic space possible—Euclidean space, Riemannian space, which may for our purposes be considered as the 3-dimensional analogue of a sphere, and a third type which is due independently to Bolyai and to Lobachevski. These three cases may be characterized by the fact that in the first a unique parallel can be drawn to a straight line from a point not lying on it, in the second none, and in the last an infinity of parallels can be drawn through the point. The theory which we are developing does not enable us to choose between these three alternatives on basis of the observations, and I therefore consider myself at liberty to restrict myself in the following to the case in which cosmic space is characterized by the Riemannian type. Our highly idealized universe is therefore one in which space is the 3-dimensional analogue of a sphere, and is accordingly finite and yet unbounded. Denoting its radius by R, the greatest distance between two points in it is πR and its total volume is $2\pi^2 R^3$. The spatial reference system introduced above is the analogue of longitude and latitude on an ordinary sphere, and the distance between two points is equal to R times the angular distance between them. But R is not necessarily independent of time as it is in some manner determined by the density of matter in the universe, and that density is decreasing in virtue of the observed expansion. In order to determine the rate at which R is increasing we recall that the distance between two nebulae is directly proportional to R and the observed rate at which this distance is increasing is such as to double it in 1,400 million years.

What is the ultimate fate of this expanding bubble of a universe? Will it continue its present rate of expansion until each island universe is completely isolated, or will it eventually cease to expand or even contract? We are here on highy speculative ground, and can only examine the various possibilities which may arise. I mentioned above that the manner in which R varies with time depends in some way on the density of the matter which constitutes the universe, and I shall now review briefly the cases which may arise if we assume the rigorous conservation of energy. It can be shown that this assumption is quite equivalent to the assumption that we are considering the material content of the universe as a gas in which the molecules (*i.e.*, nebulae!) exert no pressure on each other—and this is enough to warn us not to close our minds to other possibilities, for if, for example, we wish to take radiation into account we must also take account of the pressure which it exerts.

We first consider that highly idealized case in which the density of matter is taken as zero—the possibility which results from the assumption that the density of matter is so small as to be without effect on the structure of the universe. This case was considered fifteen years ago by the Dutch astronomer de Sitter, whose name it bears; it, together with the Einstein universe considered below, enjoys the distinction of being "stationary" in the sense that its intrinsic properties are unchanged in time. Here I must make an exception to my agreement to consider only universes in which space is of the Riemannian type, for the space of the de Sitter universe is Euclidean—its radius has become infinite. Nevertheless, each nebula is at the center of a sphere of finite radius R_d which is unchanged in time and which represents an utmost limit beyond which an observer on the nebula can not see. The other nebulae observed within his sphere appear to him to be receding with a speed which is directly proportional to their distance from him, and once they reach the limiting sphere they are forever lost to his observable universe; the space about him is expanding at such a rate that the light from nebulae outside his critical sphere is swept back again. The observations leading to the velocity-distance relation set forth above enable us to set the radius of this observable universe at 2,000 million light-years. But whatever attraction this splendid isolation may have for us, we must turn regretfully from it to other possibilities.

The remaining possibilities for a universe with zero pressure were analyzed ten years ago by the Russian mathematician Friedmann. Although his approach differed in several essential respects from that sketched in the above and although he was unable to give a satisfactory proof that no others existed, we now know that his keen analysis includes all possible cases. The fact that we can not predict precisely the fate of such a universe is due to the appearance of an arbitrary constant λ, Einstein's "cosmological constant," in the equation which determines R in terms of the density of matter. If λ is larger than a certain critical value λ_E, which is inversely proportional to the square of the total mass contained in the universe ($\lambda > \lambda_E = (\pi c^2/2kM)^2$ where c is the velocity of light, k the Newtonian constant of gravitation and M the total mass), it continues to expand without limit, so that evenually all nebulae will be lost to us. This ultimate empty state in which R has become infinite is in fact the de Sitter universe discussed above and we now see that it is stationary because everything worth happening happened long ago! On the other hand, at a finite time in the past such a monotonic universe had a zero radius. We have the choice of assuming that it began at this time as a singular point, or the emotionally more satisfactory one of assuming that our analysis breaks down because of the untenability of our hypotheses at a time when the universe was much smaller—in which case we should probably be able to conclude that it was originally shrinking and, having reached a finite lower limit, began to expand.

The next case we consider, that in which λ is less than this limiting value and yet positive ($0<\lambda<\lambda_E$), offers two possibilities depending on the magnitude of R. If R is greater than a certain critical radius $R_m>0$, which depends on λ, it continues to expand as in the case considered above, and approaches the de Sitter universe. It differs from the previous case in one essential point: at a finite time in the past it had the critical radius R_m, and if we follow it still further back we find that it was originally decreasing. The other possibility is that in which R is less than another critical value R_M, which also depends on λ and which is itself less than R_m. We then find that R increases at an eventually decreasing rate until it reaches the value R_M, at which time it begins to contract. This contraction continues until it reaches the singular state discussed above in which it has zero radius, and if we follow it through this singularity we find that it again increases, only to repeat the cycle. The case in which λ is zero (which Einstein has of late adopted) or negative is qualitatively the same as that just described; such a universe is periodic, bouncing back and forth between a state of zero radius and one in which $R = R_M$.

I have left to the last the discussion of what are undoubtedly the most interesting of universes with zero pressure, those in which λ is just equal to the critical value λ_E. In this case the two values R_m and R_M of the radius introduced above coincide, and it is possible to have a universe in which R remains at this common value R_E. This universe is the other of the stationary possibilities, and is of considerable historical interest because it is the case first considered by Einstein in 1917; and is the forerunner of all relativistic cosmologies. His general theory of relativity as first formulated did not contain the arbitrary constant λ, which he introduced two years later in order to avoid certain paradoxes associated with infinite space, and since he was interested only in stationary possibilities (the red shift indicating positive velocities of nebulae being as yet not established) he was led directly to the case which now occupies our attention. In this Einstein universe, as it is called, the radius R_E of the universe is inversely proportional to the square root of the density of the matter which it contains, or directly proportional to the total mass M. Hubble's estimate of the density of matter in nebulae (which we have expressed above in terms of the number of nebulae within the limits reached by the 100-inch) leads to the conclusion that its radius R_E is about 90,000 million light-years. This universe has the amusing property of allowing a light signal to pass clear around it—but it takes 550,000 million years! But what if R is not just equal to this critical radius R_E? Suppose it is displaced slightly from this equilibrium radius—will it return to it or deviate still further? The answer is apparent from Friedmann's work, and has recently been proved explicitly by Eddington—the equilibrium condition represented by the Einstein universe is unstable. If R is displaced by any accidental perturbation ever so little toward smaller values it continues to decrease until at the end of a finite time it becomes zero, and if it is displaced toward larger values it continues to increase and approaches the de Sitter state. This latter case, in which $\lambda = \lambda_E$ and R exceeds the critical Einstein radius R_E, has subsequently been discussed in detail by the Belgian Abbè Lemaitre. It has good claim to the special attention which it has received, for it is the only one of Friedmann's universes which has been expanding forever—no matter how far we follow it back we find that it never quite reaches the Einstein equilibrium state.[2]

With this we end our survey of the full range of possibilities for a universe in which energy is truly conserved. But allow me to repeat that we should not close our minds to the other cases which may arise. We know, for example, that in at least some portions of the universe (the interior of the stars) there is an active interplay between matter and energy; the matter is being used so as to supply radiant energy. And if Millikan's hypothesis concerning the origin of the cosmic rays proves tenable we must conclude that such an active exchange is yet more wide-spread. The effect of the annihilation of matter on the expansion of the universe has been the starting point of a series of important investigations by Tolman. This same investigator has given us an extension of the principles of thermodynamics which satisfies the fundamental criteria of the general theory of relativity

---

[2] For the benefit of those who prefer their mathematics undiluted I may point out that the exact relation to which the field equations lead is given by Friedmann's equation (Z. Physik., 10, pp. 377–396, 1922)

$$c(t - t_o) = \pm \int_{R_o}^{R} \left(\frac{x}{A - x + \lambda x^3/3}\right)^{\frac{1}{2}} dx$$

for R as a function of t, where $A = 2/3\sqrt{\lambda_E}$. For those values of R which may actually occur the cubic $P(x, \lambda)$ in x which appears in the denominator of the radical must be non-negative, and the various cases which may arise are classified according to the number and nature of the positive roots of $P(x, \lambda) = 0$. For the critical value $\lambda = \lambda_E$ the cubic has two coincident roots at $x = R_E$; R may therefore remain at this (unstable) critical radius, giving rise to the Einstein universe, or it may, among other possibilities, continually increase without limit, leading to the case considered more fully by Lemaitre. For $\lambda > \lambda_E$ we always have $P \geqq 0$, corresponding to the monotonically increasing world of the first kind. For $0 < \lambda < \lambda_E$ it has two distinct roots $R_M(\lambda) < R_m(\lambda)$ and is positive only for values of x greater than R_m or less than R_M; the former possibility leads to Friedmann's monotonic world of the second kind and the latter, in common with those cases in which $\lambda \leqq 0$ and in which P has but one real root R_M, to a periodic universe.

In a subsequent paper (Z. Physik., 21, pp. 326–332, 1924), Friedmann discussed the possibility of a world in which matter exerts no pressure but in which space is hyperbolic. The fundamental equation for this case is obtained from the above on reversing the sign of x in the cubic, and leads immediately to a monotonic increasing world of the first kind for $\lambda \geqq 0$ and a periodic world for $\lambda < 0$. A world in which space is Euclidean—the equation of which is obtained from the above by dropping the linear term in the denominator—behaves qualitatively in the same way as this hyperbolic case, a possibility which Friedmann seems to have ignored. But the existence of this possibility is apparent from the present approach, which follows an investigation by the author (Proc. Nat. Acad. Sci., 15, pp. 822–829, 1929) in which all possibilities are explicitly indicated. The special monotonic case in which $\lambda = 0$ has more recently been considered in detail by Einstein and de Sitter (Proc. Nat. Acad. Sci., 18, pp. 213–214, 1932).

Although we have for the sake of simplicity restricted ourselves to a discussion of Friedmann's worlds in which energy is rigorously conserved, the qualitative discussion of types thus found will hold under the much weaker assumption that the energy and the pressure are both positive non-increasing functions of R; a detailed analysis of all such possibilities is included in a comprehensive report on the subject which is to appear soon in Rev. Mod. Physics.

and has applied it to the question of the thermodynamic relations exhibited by the universes contemplated in relativistic cosmology. He has, for example, thus sought to establish the possibility of a universe in which radiation is in equilibrium with matter and which, although expanding or contracting at a finite rate, does not suffer the ultimate "heat death" which an observer viewing it through the eyes of classical thermodynamics would predict.

Finally, I must call your attention to a doubt which the results outlined above have raised. The time scale which the observed red shift in light from the distant nebulae leads to if interpreted as due to velocity is rather meager. What are we to think of a universe whose radius is at present expanding at such a rate as to double itself every 1,400 million years which contains stars whose age is estimated at millions of millions of years?[3] Perhaps we may be able to conclude that the processes which lead us to these tremendous ages were proceeding at a much greater rate when the world was young, or it may be that the astronomers have been over-zealous in demanding millions of millions of years when but a fraction of that would have sufficed. Einstein and de Sitter appear to have been moved by the rather short time scale to favor a periodic universe in which we are now enjoying the expansion phase, but which may conceivably reverse this tendency before the sun becomes too cold to support life. In addition to those who believe that the at first sight paradoxical time scale is nevertheless reconcilable with the observed facts there exists a group which would attribute the observed red shift, which we have throughout interpreted as a velocity, to a property of light which has traveled the tremendous inter-nebular distances. Zwicky suggested a few years ago that there may exist a mechanism by which the light-corpuscles surrender a minute fraction of their energy to nebulae and other matter which they pass on their journey to us; this loss of energy would be proportional to the distance through which they travel and would, in accordance with our present theory of light, give rise to a red shift in the observed spectrum. In this case our interpretation would be quite false—the observed red shift would be due to the properties of "tired" light rather than to the nebulae themselves. But I do not believe that even if room could be found in our theories for such a modification it would alter essentially the general outlook with which we have been concerned this evening, for so long as we have sufficient evidence in other fields to hold to the general theory of relativity and so long as the homogeneity assumptions with which we started are not at variance with the observations we may consider relativistic cosmology as a simple corollary of the relativity theory. Robbed of all contact with the empirical we would of course be unable to decide which of the alternatives was best suited for a description of the actual universe—perhaps we should fall back on the Einstein universe which was originally offered to us as escape from the paradoxes of an infinite world filled uniformly with matter. But in the lack of further facts I should prefer to wield Occam's razor on all *ad hoc* explanations of the red shift and accept that one which follows so naturally from our present views of the nature of the physical world, the bold outlines of which I have had the pleasure of sketching before you this evening.

[3] I here refer to the attitude which has been expressed by de Sitter (Bull. Astron. Inst. Netherlands 5 No. 193, p. 212, 1930) and which has been adopted by others. I do not consider the objection to be as serious, but do hold it to be a valid argument for a universe of the type $\lambda = \lambda_E$ resulting from a perturbation of the unstable Einstein world.

OBITUARY

CHARLES WILLISON JOHNSON

It has been said of many great men that kindliness of manner and disinterested helpfulness were among their outstanding traits. Joseph Leidy is remembered by those he taught almost as much for these qualities as for the greatness of his intellect or his innumerable and far-reaching discoveries. Into the early lives of many of us standing awestruck at the threshold of the world of nature, which we wished so much to know better, a hand was stretched out, and a kindly teacher—or better, friend—led our faltering steps through the portal and fixed our life's greatest interest. This brief tribute is to one whose guiding hand placed me on the happy road which teaches boys to see, to understand and to appreciate the world about them.

Charles Willison Johnson was born at Morris Plains, Morris County, New Jersey, on October 26, 1863. Educated in public and private schools at Morristown, New Jersey, he early showed a deep interest in natural history. In 1881, his family removed to St. Augustine, Florida, and there during the succeeding seven years he continued his studies and made extensive collections, chiefly of insects, mollusks and fossils.

In 1888, having been appointed curator of the Museum of the Wagner Free Institute of Science in Philadelphia, he brought to this work a broad knowledge of natural history and an intimate acquaintance with the existing and fossil fauna of Florida. At