# TWO METHODS OF INVESTIGATING THE NATURE OF THE NEBULAR RED-SHIFT*

EDWIN HUBBLE AND RICHARD C. TOLMAN[1]

ABSTRACT

Two theoretically possible methods of investigating the nature of the nebular red-shift have been known for some time. These two methods depend on the relations connecting either nebular dimensions with observed luminosities or nebular counts with observed luminosities, which might be expected to hold under the alternative assumptions of mutual recession or some other cause for the red-shift. The main purpose of the present article is to study in detail the procedure that must be adopted in applying these methods to actual observations in view of such factors as the indefiniteness of nebular diameters resulting from their gradual decrease in surface brightness from the center outward, the necessity for relating apparent bolometric luminosities to observed photographic magnitudes, and the effect on nebular counts produced by a wide spread in the values of absolute nebular magnitudes.

In addition to this study, preliminary statements are made as to the present status of the observational findings. In the case of the relation between nebular dimensions and luminosities, the observations are such as to confirm our general ideas as to the extra-galactic character of the objects in question, but are not yet sufficient to permit a decision between recessional or other causes for the red-shift. In the case of the relation between nebular counts and luminosities, the observations now available show a rate of increase in counts with distance which seems rather large compared with what would be expected for a homogeneous distribution of nebulae on the basis of either a recessional or a non-recessional theory. Nevertheless, by assuming no appreciable absorption of light on the path from the nebulae and by ascribing a rather high effective temperature to the nebulae, it might be possible to explain the counts on the basis of either a static homogeneous model with some unknown cause for the red-shift or an expanding homogeneous model having a positive spatial curvature which introduces effects that seem unexpectedly large but may not be impossible. This conclusion must be considered as tentative, however, and may need serious modification as a result of further work.

## The Nature of the Problem

### 1. INTRODUCTION

Light arriving from the extra-galactic nebulae exhibits a shift toward the red in the position of its spectral lines which is approximately proportional to the distance to the emitting nebula. The most obvious explanation of this finding is to regard it as directly correlated with a recessional motion of the nebulae, and this assumption has been commonly adopted in the extensive treatments of nebular motion that have been made with the help of the relativistic theory of gravitation, and also in the more purely kinematical treat-

* *Contributions from the Mount Wilson Observatory, Carnegie Institution of Washington,* No. 527.

[1] Of the California Institute of Technology.

ment proposed by Milne. Nevertheless, the possibility that the red-shift may be due to some other cause, connected with the long time or distance involved in the passage of light from nebula to observer, should not be prematurely neglected; and several investigators have indeed suggested such other causes, although without as yet giving an entirely satisfactory detailed account of their mechanism.

Until further evidence is available, both the present writers wish to express an open mind with respect to the ultimately most satisfactory explanation of the nebular red-shift and, in the presentation of purely observational findings, to continue to use the phrase "apparent" velocity of recession. They both incline to the opinion, however, that if the red-shift is not due to recessional motion, its explanation will probably involve some quite new physical principles.

## 2. PLAN OF TREATMENT

The purpose of the present article is to give detailed consideration to two methods of investigating the nature of the nebular red-shift which, in principle, could be used to distinguish between recessional motion and at least some of the other explanations that might be offered to account for the shift. The first of these methods depends on the relations between nebular dimensions and observed luminosities which would be expected for a homogeneous distribution of nebulae under the alternative assumptions of mutual recession or of some other cause for the red-shift; the second method depends on the relations between nebular counts and observed luminosities which would be expected under those same circumstances. The general theoretical nature of the two methods has been evident for some time, and the main function of the present article is the practical one of investigating the complexities introduced into the treatment of actual observations by such factors as the indefiniteness of nebular diameters connected with their gradual decrease in surface brightness from the center outward, the necessity for relating theoretical luminosities to observed photographic magnitudes, and the effect on nebular counts produced by a wide spread in the actual values of absolute nebular magnitudes.

Part I of the article presents relations which might be expected to hold in case the nebular red-shift is due to recession. In treating the problem the actual universe is represented by a homogeneous expanding model obeying the relativistic laws of gravitation. This choice of model is recommended by several considerations. The model is governed by the principles of general relativity, which are known to furnish the most adequate description of gravitational phenomena now available. These principles, moreover, are quite definite and provide a sufficient basis for calculating the relations needed for the proposed tests of the nature of the red-shift. The general type of model is familiar, and, by the assignment of suitable values to the different parameters involved in the equations describing the type, a wide variety of special cases becomes available for comparison with observation. Furthermore, it has been shown by the work of Robertson, of Kermack and McRae, and of Milne himself that one of these special cases leads to important observational properties which are approximately the same as would be predicted from Milne's kinematic model—a circumstance which makes it less necessary to give a separate treatment to that model.

Part II of the article presents similar relations for the case that the nebular red-shift is due to causes other than recession. In treating this second problem—prior to any satisfactory detailed theory of such other causes for the red-shift—it is of course not very clear just what model should be taken as a representation of the actual universe. For definiteness we actually use a static Einstein model of the universe, combined with the assumption that the photons emitted by a nebula lose energy on their journey to the observer by some unknown effect, which is linear with distance and which leads to a decrease in frequency without appreciable transverse deflection and, in particular, without any decrease in rate of arrival at the observer. As an objection to the use of this picture it might be urged that the Einstein model is known to be unstable and subject to expansion or contraction if disturbed; and, indeed, the improbability of any kind of mechanics which would permit a stable static distribution of nebulae might also be stressed. Our choice of model is, nevertheless, actually adapted to our present interests since we wish to investigate the possibility that the main cause for the red-shift may not be recession; and we may hence properly use the Einstein static universe as a limiting case of models in which the effect of expansion or contraction would be negligible compared with that due to the actual (unknown) cause of the red-shift.

Finally, in Part III we make a preliminary statement as to the present status of observational work.

## Part I. Treatment Assuming Red-Shift Due to Recession

### 3. THE THEORETICAL FORMULAE

The present section gives certain theoretical formulae which have already been derived as applying to an expanding cosmological model on the basis of relativistic mechanics,[2] while the next section considers the modifications necessary to adapt the formulae for use in connection with actual observations.

For the line element of the model we may take the expression

$$ds^2 = -e^{g(t)}\left(\frac{d\bar{r}^2}{1-\bar{r}^2/R_0^2}+\bar{r}^2 d\theta^2+\bar{r}^2 \sin^2\theta d\phi^2\right)+dt^2 , \tag{1}$$

where $R_0$ is a constant parameter related to the pressure and density of material in the model, and the bar over the radial co-ordinate $\bar{r}$ is used to distinguish it from a different radial co-ordinate, $r$, which is perhaps more often employed but which is less convenient for our immediate purposes. The spacelike co-ordinates $\bar{r}$, $\theta$, and $\phi$, in this form of writing the line element, are co-moving in character, so that except for small peculiar velocities the nebulae have no motion relative to these co-ordinates, and the dependence of the model on time is introduced by the appearance of the as yet unspecified function $g(t)$. The timelike co-ordinate $t$ is such as to agree with ordinary measurements of proper time as made by any local observer at rest with respect to $\bar{r}$, $\theta$, and $\phi$.

In accordance with this form of line element, the red-shift in the wave-length of light arriving from a distant nebula is given by the formula

$$\frac{\lambda+\delta\lambda}{\lambda}=e^{\frac{1}{2}(g_2-g_1)} , \tag{2}$$

[2] The derivation of these formulae in the same notation as that used here will be found in Tolman's *Relativity, Thermodynamics and Cosmology*, Oxford, 1934. The content of equations (1), (2), and (5) is well known. The original derivation of equations (3) and (4) was given by Tolman, *Proc. Nat. Acad.*, **16**, 511, 1930.

where $g_1$ is the value of $g(t)$ at the time the light observed leaves the nebula, and $g_2$ the value of that function at the time of arrival at the observer, both source and observer being regarded as substantially at rest with respect to $\bar{r}$, $\theta$, and $\phi$.

For the angular diameters of essentially similar nebulae located at different values of the radial co-ordinate $\bar{r}$, but simultaneously observed at the origin $\bar{r}=0$, we have

$$\delta\theta = \frac{\text{const.}}{\bar{r}}\left(\frac{\lambda+\delta\lambda}{\lambda}\right) , \tag{3}$$

where $\delta\lambda/\lambda$ is the observed red-shift for the nebula in question.

For the bolometric luminosity of such similar nebulae, as measured outside the earth's atmosphere in ergs per square centimeter per second, we have

$$l_b = \frac{\text{const.}}{\bar{r}^2}\left(\frac{\lambda}{\lambda+\delta\lambda}\right)^2 . \tag{4}$$

And finally for the number of nebulae in a given co-ordinate range $d\bar{r}$, we have

$$dN = \text{const.}\ \frac{\bar{r}^2 d\bar{r}}{(1-\bar{r}^2/R_0^2)^{\frac{1}{2}}} . \tag{5}$$

### 4. RE-EXPRESSION OF FORMULAE FOR USE WITH OBSERVATIONS

We now turn to the re-expression of the foregoing formulae in forms adapted for application to the discussion of observational findings.

*a) Empirical form for the function* $g(t)$.—Actual observations on the nebulae have not yet reached a stage which will permit a decision between the different theoretically possible forms for the function $g(t)$ that have so far been investigated. Enough progress has been made, however, to show that this function should be taken as approximately linear in $t$ over a long time-interval in the neighborhood of the present. Hence we shall express the function in the form of a power series,

$$g(t) = 2(kt + lt^2 + mt^3 + \cdots) , \tag{6}$$

developed around the time of observation $t_2=0$, which is taken for convenience as the starting-point for temporal measurements. The factor 2 in this expression has been introduced to avoid the later appearance of fractions, and the quantities $k, l, m \ldots .$, are to be regarded as empirical constants, there being no loss in generality in omitting a constant term in $t^0$.

*b) Red-shift as a function of co-ordinate position.*—Substituting (6) into our original expression for the red-shift (2), we may now write

$$z=\frac{\delta\lambda}{\lambda}=e^{-kt_1-lt_1^2-mt_1^3-\cdots}-1\,, \tag{7}$$

where $t_1$ is the time of departure of light from the nebula under consideration, and the symbol $z$ has been introduced as an alternative expression for $\delta\lambda/\lambda$ to simplify the writing of certain later formulae.

This expression relates the red-shift to the time of departure of light from the source $t_1$, but can readily be changed into forms relating red-shift to the co-ordinate position of the source $\bar{r}$ with the observer at the origin. Setting $ds$ equal to zero in our expression for the line element (1), we can obtain an expression for the co-ordinate velocity of light and thus connect the time of departure $t_1$ with co-ordinate position $\bar{r}$ by the equation

$$\int_0^{\bar{r}}\frac{d\bar{r}}{(1-\bar{r}^2/R_0^2)^{\frac{1}{2}}}=\int_{t_1}^{0}e^{-kt-lt^2-mt^3-\cdots}dt\,, \tag{8}$$

which can be shown to lead to a series expression for red-shift in terms of co-ordinate position of the emitting nebula of the form

$$z=\frac{\delta\lambda}{\lambda}=k\bar{r}-l\bar{r}^2+\left(\frac{k}{6R_0^2}+\tfrac{1}{3}kl+m\right)\bar{r}^3-\cdots\,. \tag{9}$$

Observations on red-shift as a function of nebular magnitudes have not yet progressed far enough to determine the coefficients in the foregoing expression beyond the first one, $k$, which, however, is known with considerable accuracy. On the other hand, since magnitudes can be determined for objects too faint for spectroscopy, it has already become necessary in treating the data on nebular counts to a given limiting magnitude to extrapolate the expression for red-shift beyond the range over which it has been measured. Since even the signs of $l$ and $m$ are dubious, we shall for the present make the necessary extrapolation by treating these constants as equal to zero. Equation (8) combined with (7) then gives

$$z=\frac{\delta\lambda}{\lambda}=kR_0 \sin^{-1}\frac{\bar{r}}{R_0}=kR_0 \sin^{-1} x \tag{10}$$

as the desired expression for red-shift in terms of co-ordinate position, where we use $x$ as a convenient abbreviation for the ratio $\bar{r}/R_0$.

*c) Observed radii for corresponding points on similar nebulae.*—The expression for observed angular diameter as a function of co-ordinate position and red-shift given by (3) would be immediately applicable to spherical nebulae having a sharp cut-off in surface brightness at a definite radius. Actual nebulae, however, are found to give photographic images that continue to increase in size with time of exposure, owing mainly to a gradual fading-out of surface brightness as we proceed from the center of the nebula outward. For this reason equation (3) is not directly applicable and is most conveniently replaced by the evident generalization

$$\frac{\rho'}{\rho}=\frac{\bar{r}}{\bar{r}'}\frac{(1+z')}{(1+z)}, \tag{11}$$

where $\rho'$ and $\rho$ are the observed values of the radii for corresponding points on two similar nebulae, located at $\bar{r}'$ and $\bar{r}$ and giving the observed red-shifts $z'$ and $z$. As corresponding points on the two nebulae we may take points along specified pairs of similar axes at which the surface brightness of the nebulae has dropped to some given fraction of its central value.

*d) The relation between co-ordinate position and photographic magnitude.*—The relation given by (4) connecting the co-ordinate position and red-shift of similar nebulae with their bolometric luminosities as measured outside the earth's atmosphere is not immediately applicable, since in actual practice the luminosities of nebulae are determined by comparing their photographic magnitudes with those for stars in selected calibrated areas.

To relate photographic magnitudes to bolometric luminosities, we may first define the bolometric magnitude of an object $m_b$ in terms of its bolometric luminosity $l_b$, and its photographic magnitude $m_p$ in terms of its photographic luminosity $l_p$, by the equations

$$m_b = 2.5 \log \frac{1}{l_b} + \text{const.} , \tag{12}$$

$$m_p = 2.5 \log \frac{1}{l_p} + \text{const.} , \tag{13}$$

where the additive constants are introduced to permit whatever conventions are desired as to starting-points for the magnitude scales and where the logarithms are to the base 10.

The quantity $l_b$, called bolometric luminosity, occurring in the first of these equations, has already been defined as the energy flux in ergs per square centimeter per second arriving from the object in question outside the earth's atmosphere, and we may write therefor

$$l_b = \int_0^\infty I(\lambda) d\lambda , \tag{14}$$

where $I(\lambda)d\lambda$ is the energy flux arriving between wave-lengths $\lambda$ and $\lambda + d\lambda$.

The quantity $l_p$, called "photographic luminosity," occurring in equation (13), is conceptually less satisfactory than $l_b$, owing to the complicated nature of photographic processes. Both in the calibration and in the measurement of stellar magnitudes, however, the attempt is made to treat $l_p$ as a definite quantity, which can be obtained by multiplying the energy flux received from the object in question in each wave-length range, $d\lambda$, by a factor proportional to the photographic effectiveness of unit flux at that wave-length and then integrating over all wave-lengths. Hence, if $I(\lambda)$ again refers to the flux outside the earth's atmosphere and $a(\lambda)$ is the fractional transmission through the earth's atmosphere and the instrument in use, we may write

$$l_p = \int_0^\infty I(\lambda) a(\lambda) p(\lambda) d\lambda , \tag{15}$$

where $p(\lambda)$ expresses the photographic effectiveness of unit energy flux at the wave-length $\lambda$. We thus obtain, at least formally, a definite expression for $l_p$, and hence also for $m_p$, which is defined in terms thereof.[3]

It will be convenient to introduce the symbol $p$ for the ratio of photographic to bolometric luminosity,

$$p=\frac{l_p}{l_b}=\frac{\int_0^\infty I(\lambda)a(\lambda)p(\lambda)d\lambda}{\int_0^\infty I(\lambda)d\lambda} . \tag{16}$$

Subtracting (12) from (13) and introducing (14), (15), and (16), we can then also write

$$\Delta m_{pb}=m_p-m_b=\text{const.}-2.5\log p \tag{17}$$

as an expression for the difference between the photographic and bolometric magnitudes of a celestial object, where the additive constant depends merely on the starting-points selected for the magnitude scales.

We are now ready to consider the relation between co-ordinate position and photographic magnitude. Substituting (16) into (4),

[3] The most serious difficulties connected with the concept of photographic luminosity, as defined by eq. (15), arise from the fact that the values assigned to the luminosities of objects must be determined from the photographic effects which they produce, and such effects depend in general in a complicated, non-linear manner on many factors. The following difficulties may be mentioned: Photographic effects may be estimated in a variety of ways, e.g., from the diameter of a stellar image, the exposure necessary for a threshold image, the density of schraffierkassette images, etc. The extent of such effects is not in general a linear function of intensity of illumination and time of exposure. The relative effects at different wave-lengths depend on the type of plate used, on the total strength of the exposure, and on the method of development. Little is known about the combined effect resulting from the superposition of light of different wave-lengths.

These difficulties are to some extent met as follows: The results of different methods of estimating photographic effects are empirically correlated. Pairs of objects in selected areas of the sky are calibrated as to relative luminosity with the help of neutral screens which make their photographic effects agree; standards of luminosity are thus obtained with which other objects may be compared. Photographic luminosities are regarded as defined with respect to a given type of plate and instrument. The values assigned to $p(\lambda)$ are determined from monochromatic exposures which lead to a density of image comparable with that in astronomical work, when developed in a standard manner.

we obtain for the photographic luminosity of a given nebula as a function of its co-ordinate position $\bar{r}$ and fractional red-shift $z$ the expression

$$l_p=\frac{\text{const.}}{\bar{r}^2}(1+z)^{-2}p \ . \tag{18}$$

Combining with the definition for photographic magnitude given by (13) and taking common logarithms, we can write (18) in the form

$$\log \bar{r}=0.2\{m-5\log(1+z)+2.5\log p\}+\text{const.} \tag{19}$$

Furthermore, for convenience, we may choose our scale for the radial co-ordinate in such a way that $\bar{r}=10$ for an object located at the present time at the usual standard distance of 10 parsecs. As a special case of (19) we may then write

$$\log 10=0.2(M+2.5\log P)+\text{const.}, \tag{20}$$

where $M$ is the absolute magnitude of the nebula in question, the term involving the red-shift has become negligible for this now near-by object, and $P$ is the value given by equation (16) with the spectral distribution which would be shown by the nebula in the absence of red-shift. Subtracting (20) from (19), we then finally obtain, as the desired expression connecting the co-ordinate position of a nebula $\bar{r}$ with its measured photographic magnitude $m$, absolute photographic magnitude $M$, and fractional red-shift $z$,

$$\log \bar{r}=0.2\{m-5\log(1+z)-K-M\}+1 \ , \tag{21}$$

where the correction term $K$ is given by

$$K=2.5\log\frac{P}{p} \ . \tag{22}$$

Two somewhat different methods for evaluating $K$ are available. The first[4] is a semi-empirical method which relates $K$ to the accepted corrections for passing from bolometric to photographic magnitudes for stars of known spectral class; the second[5] is a more computational

[4] Hubble and Humason, *Mt. W. Contr.*, No. 427; *Ap. J.*, **74**, 43, 1931.

[5] W. de Sitter, *Bull. Astron. Inst. Netherlands*, **7**, 205, 1934.

one, involving an integration over the sensitive range of the photographic plate.

To make use of the semi-empirical method, we first note that equations (17) and (22) can be combined to give

$$K=\Delta m_{pb}-\Delta M_{pb}\ , \tag{23}$$

where $\Delta M_{pb}$ is the difference between the photographic and bolometric magnitudes of the nebula which would be found in the absence of any red-shift, and $\Delta m_{pb}$ is this difference in the presence of the actually existing red-shift. These differences are, of course, immediately related to the unshifted and the shifted spectral distributions in the radiation emitted by the nebula.

From an examination of nebular spectra it proves possible, at least within limits, to identify the unshifted distribution as approximately similar to that from stars of a definite spectral class. Furthermore, making use of the known fairly satisfactory correlation between spectral classes and temperatures, we can regard this unshifted distribution as equivalent to that emitted by a black body at a definite temperature $T_0$, and take the effect of a superimposed uniform red-shift $\delta\lambda/\lambda=z$ as leading to a shifted distribution such as would be emitted by a source at the lower temperature $T_0(1+z)^{-1}$. Hence it is possible also to identify the shifted distribution as approximately similar to that from stars of a definite later spectral class. We may therefore calculate values of $K$ for any selected initial distribution of radiation and red-shift $z$ by substituting into (23) the accepted values of $\Delta M_{pb}$ and $\Delta m_{pb}$ that have been determined for stars of the spectral classes involved.

The method just outlined may be regarded as semi-empirical, since for stars of any given spectral class we may use the relation

$$\Delta m_{pb}=CI+HI+\Delta m_r\ , \tag{24}$$

where we can take the color index $CI$ as the actual empirical difference between the photographic and visual magnitudes of such stars, the heat index $HI$ as the empirical difference between their visual and radiometric magnitudes, and the correction to no atmosphere $\Delta m_r$ as computed from empirical transmission coefficients.

To use the more purely computational method for evaluating the correction $K$, we may start by combining equations (16) and (22) to give

$$K=2.5 \log \frac{\int_0^\infty I(\lambda)d\lambda}{\int_0^\infty I_0(\lambda_0)d\lambda_0} \times \frac{\int_0^\infty I_0(\lambda_0)a(\lambda_0)p(\lambda_0)d\lambda_0}{\int_0^\infty I(\lambda)a(\lambda)p(\lambda)d\lambda}, \qquad (25)$$

where $I(\lambda)\, d\lambda$ is the energy flux in the wave-length range $\lambda$ to $\lambda+d\lambda$ arriving from the actual nebula located at $\bar{r}=\bar{r}$, and $I_0(\lambda_0)d\lambda_0$ is the flux that would be arriving in the range $\lambda_0$ to $\lambda_0+d\lambda_0$ from a hypothetical similar nebula at the standard distance where $r=10$.

To obtain actual figures with the help of this expression, we may treat the unshifted distribution $I_0(\lambda_0)$ as corresponding to that from a black body at temperature $T_0$, and the shifted distribution $I(\lambda)$ as corresponding to the lower temperature $T_0(1+z)^{-1}$, where $\delta\lambda/\lambda=z$ is the fractional red-shift, which, in accordance with the Planck distribution law, gives us

$$K=2.5 \log \frac{\int_0^\infty \lambda^{-5}\left(e^{\frac{c_2(1+z)}{\lambda T_0}}-1\right)^{-1} d\lambda}{\int_0^\infty \lambda^{-5}\left(e^{\frac{c_2}{\lambda T_0}}-1\right)^{-1} d\lambda} \times \frac{\int_0^\infty \lambda^{-5}\left(e^{\frac{c_2}{\lambda T_0}}-1\right)^{-1} a(\lambda)p(\lambda)d\lambda}{\int_0^\infty \lambda^{-5}\left(e^{\frac{c_2(1+z)}{\lambda T_0}}-1\right)^{-1} a(\lambda)p(\lambda)d\lambda}. \qquad (26)$$

Here the value of the radiation constant $c_2$ is known, and it has not been necessary to consider the absolute values of $I$ and $I_0$, owing to their double occurrence in both numerator and denominator. This expression is easily reduced to

$$K=2.5 \log \frac{\int_0^\infty \lambda^{-5}\left(e^{\frac{c_2}{\lambda T_0}}-1\right)^{-1} a(\lambda)p(\lambda)d\lambda}{(1+z)^4\int_0^\infty \lambda^{-5}\left(e^{\frac{c_2(1+z)}{\lambda T_0}}-1\right)^{-1} a(\lambda)p(\lambda)d\lambda}. \qquad (27)$$

Referring to (21), we can also, if desired, combine the two corrections to be subtracted from the photographic magnitude $m$ into the form

$$\Delta m = 5 \log (1+z) + K$$

$$= 2.5 \log \frac{\int_0^\infty \lambda^{-5}\left(e^{\frac{c_2}{\lambda T_0}} - 1\right)^{-1} a(\lambda)p(\lambda)d\lambda}{(1+z)^2 \int_0^\infty \lambda^{-5}\left(e^{\frac{c_2(1+z)}{\lambda T_0}} - 1\right)^{-1} a(\lambda)p(\lambda)d\lambda} . \qquad (28)$$

It should be specially noted that this expression differs from the correction to $m$ proposed by de Sitter,[6] which contains the term $(1+z)^3$ instead of $(1+z)^2$. Expression (28), however, would seem to give the proper correction to use in connection with our equation (21), since it has been derived in such a way as to make appropriate allowance, first, for the double effect of nebular recession in reducing both the individual energy and the rate of arrival of photons, and then for the further circumstance that a change in spectral distribution of the energy that does arrive will lead to changes in its photographic effectiveness. It is to be noted that the necessity of allowing for the double effect of recession, both on the individual energy and the rate of arrival of photons, is analogous to the necessity of using the quadratic factor

$$[\sqrt{1 - V^2/c^2}/(1+V/c)]^2 = (1+z)^{-2}$$

in making the ordinary Lorentz transformation for the component of a Poynting vector lying in the direction of the relative velocity $V$ of the two systems of co-ordinates considered.

In concluding this section it may be mentioned that we have already used both the semi-empirical and the computational methods to obtain values of $K$ as a function of the red-shift $\delta\lambda/\lambda = z$, for an assumed range of values for the spectral class and temperature of the nebulae without red-shift, and have found reasonable agreement between the results obtained by the two methods. This now completes the long discussion necessary to replace the simple theoretical relation (4) between co-ordinate position and luminosity, by the prac-

[6] *Loc. cit.*, equations (26), (27), and (28).

tical relation (21) between co-ordinate position and photographic magnitude with its bothersome but necessary correction term $K$.

*e) Allowance for spread in absolute nebular magnitudes.*—As the final step in translating the theoretical formulae of section 3 into their practical form, we shall find it desirable to modify formula (5) for the number of nebulae in a given co-ordinate range in such a way as to recognize the actual spread in absolute nebular magnitudes. Since this spread is observationally found[4,7] to be approximately a Gaussian distribution with a certain dispersion $\sigma$ around a mean absolute magnitude $M_0$, we can re-write (5) in the form

$$dN=\frac{C}{\sqrt{2\pi}\sigma}e^{-\frac{(M-M_0)^2}{2\sigma^2}}\frac{\bar{r}^2d\bar{r}}{(1-\bar{r}^2/R_0^2)^{\frac{1}{2}}}dM\ , \tag{29}$$

where $C$ is a constant and $dN$ now becomes the number of nebulae located in the co-ordinate range $\bar{r}$ to $\bar{r}+d\bar{r}$ and having absolute magnitudes between $M$ and $M+dM$.

## 5. DISTRIBUTION OF LUMINOSITY OVER THE SURFACE OF NEBULAR IMAGES

After the foregoing long preliminary we are now ready to consider the use of our formulae to obtain tests of the nature of the red-shift. Any such test must be expressible in language which eliminates the co-ordinate $\bar{r}$ as the measure of nebular distance, since this is an auxiliary rather than a directly observable quantity. As a first method of eliminating $\bar{r}$, we can combine equations (3) and (4) to give

$$\frac{\delta\theta}{l_b^{\frac{1}{2}}}=\text{const.}\left(\frac{\lambda+\delta\lambda}{\lambda}\right)^2 \tag{30}$$

as a connection between the observable quantities—angular diameter, bolometric luminosity, and red-shift—for the case of intrinsically similar nebulae.[8] For practical purposes, however, this simple formula must be replaced by expressions which allow for the use of photographic methods and for the absence of sharp nebular boundaries.

[7] Hubble and Humason, *Mt. W. Comm.*, No. 116; *Proc. Nat. Acad.*, **20**, 264, 1934.

[8] Tolman, *Proc. Nat. Acad.*, **16**, 511, 1930.

Consider two elliptical nebulae, located at $\bar{r}$ and $\bar{r}'$, but similar in intrinsic character and so oriented as to present similar aspects to an observer at the origin, and let $i$ and $i'$ be the intensities of photographic illumination for corresponding points on their photographic images. Since the total photographic luminosity of an object is proportional to the integrated intensity of illumination over its image, and since for elliptical nebulae the character of the spectral distribution is approximately independent of position on the luminous surface, we can put

$$\frac{idS}{i'dS'}=\frac{l_p}{l'_p}, \tag{31}$$

where $dS$ and $dS'$ are corresponding elements of area on the two images, and $l_p$ and $l'_p$ are the total photographic luminosities.

In accordance with the expression for the observed radii of corresponding points given by (11), we can write

$$\frac{dS}{dS'}=\frac{\bar{r}'^2}{\bar{r}^2}\frac{(1+z)^2}{(1+z')^2}, \tag{32}$$

while from (18)

$$\frac{l_p}{l'_p}=\frac{\bar{r}'^2(1+z')^2 p}{\bar{r}^2(1+z)^2 p'}. \tag{33}$$

Substituting (32) and (33) in (31), we obtain

$$\frac{i}{i'}=\frac{(1+z')^4}{(1+z)^4}\frac{p}{p'} \tag{34}$$

as an expression for the ratio of the intensities of photographic illumination at corresponding points on the two nebular images.

If desired, we can introduce into (34) our present knowledge of the distribution of photographic intensity over the surface of images of elliptical nebulae by using the empirical equation[9]

$$i=\frac{i_0}{\left(\frac{\rho}{a}+1\right)^2}, \tag{35}$$

[9] Hubble, *Mt. W. Contr.*, No. 398; *Ap. J.*, 71, 231, 1930.

which gives a good expression for the relation between the photographic intensity $i$ at the radius $\rho$ from the center of the elliptical image, in terms of the central intensity $i_0$ and a parameter $a$, which, except in the case of circular images, depends on the direction of the radial axis.

Assuming that relation (35) holds for a nebula at $\bar{r}'$ and using the relation between $\rho$ and $\rho'$ given by (11), we can then re-write (34) for a similar nebula at $\bar{r}$ in the form

$$i=\frac{i_0' \dfrac{(1+z')^4}{(1+z)^4}\dfrac{p}{p'}}{\left\{\dfrac{\rho}{a'}\dfrac{\bar{r}(1+z')}{\bar{r}'(1+z)}+1\right\}^2}\,, \tag{36}$$

so that we may now take

$$i=\frac{i_0}{\left(\dfrac{\rho}{a}+1\right)^2}\,, \tag{37}$$

with

$$i_0=\frac{\text{const.}}{(1+z)^4}\,p \qquad \text{and} \qquad a=\frac{\text{const.}}{\bar{r}}\,(1+z)\,, \tag{38}$$

as a general expression for the distribution of surface intensity over the images of similar nebulae at different co-ordinate positions. With the aid of (21) and (22), the expressions (38) for $i_0$ and $a$ can also be written in the form

$$\left.\begin{aligned}\log i_0&=\text{const.}-4\log(1+z)-0.4K\,,\\ \log a&=\text{const.}-0.2m+2\log(1+z)+0.2K\,.\end{aligned}\right\} \tag{39}$$

Since the red-shift $z$ and the photographic magnitude $m$ are directly measurable quantities and the correction term $K$ can be calculated from the spectral distribution of nebular radiation, these results provide a possibility for testing the assumption of nebular recession.

### 6. NEBULAR COUNTS TO A GIVEN LIMITING MAGNITUDE

We now turn to a second method of eliminating the co-ordinate $\bar{r}$ so as to obtain a test of our model. This is made possible by the pair of relations which give photographic magnitudes and numbers of nebulae as functions of $\bar{r}$.

In accordance with (29) we have

$$dN=\frac{C}{\sqrt{2\pi}\sigma}\, e^{-\frac{(M-M_0)^2}{2\sigma^2}}\,\frac{\bar{r}^2 d\bar{r}}{(1-\bar{r}^2/R_0^2)^{\frac{1}{2}}}\, dM \tag{40}$$

for the number of nebulae in the co-ordinate range $\bar{r}$ to $\bar{r}+d\bar{r}$, which have an absolute magnitude lying in the range $M$ to $M+dM$. Integrating this expression from $\bar{r}=0$ out to a particular value $\bar{r}=\bar{r}_1$, we obtain for the number of nebulae with the specified absolute magnitude which lie inside of $\bar{r}_1$

$$dN=\frac{C}{3\sqrt{2\pi}\sigma}\,\bar{r}_1^3\left[\frac{3R_0^3}{2\bar{r}_1^3}\left(\sin^{-1}\frac{\bar{r}_1}{R_0}-\frac{\bar{r}_1}{R_0}\sqrt{1-\frac{\bar{r}_1^2}{R_0^2}}\right)\right]e^{-\frac{(M-M_0)^2}{2\sigma^2}}\,dM\,. \tag{41}$$

For small values of $\bar{r}_1/R_0$ this expression will make the number of nebulae in question nearly proportional to $\bar{r}_1^3$, since the quantity inside the square brackets will then be nearly unity. For increasing values of $\bar{r}_1/R_0$, however, the effects of curvature will become more and more important.

In order to obtain from equation (41) information as to the number of nebulae brighter than a given apparent magnitude $m_1$, instead of the number out to a given value of the radial co-ordinate $\bar{r}_1$, we may use relation (21) connecting $m$ and $\bar{r}$. For the purposes of our further treatment it will be advantageous to make use of (21) at once to replace the factor $\bar{r}_1^3$ outside the brackets by the corresponding expression in $m_1$; inside the brackets, however, it will prove convenient for the present to retain $\bar{r}_1$, substituting, for simplicity, our previous notation

$$x=\bar{r}/R_0\,. \tag{42}$$

Carrying this out, we then obtain

$$\left.\begin{aligned}dN=&\frac{1000C}{3\sqrt{2\pi}\sigma}\,10^{0.6[m_1-5\log(1+z_1)-K_1-M]}\\&\times\left[\frac{3}{2x_1^3}\left(\sin^{-1}x_1-x_1\sqrt{1-x_1^2}\right)\right]e^{-\frac{(M-M_0)^2}{2\sigma^2}}dM\,,\end{aligned}\right\} \tag{43}$$

which, by a simple transformation, can be written in the form

$$dN=\frac{1000C}{3\sqrt{2\pi\sigma}}\,10^{-0.6(M_0-0.6909\sigma^2)}10^{0.6m_1}10^{-0.6[5\log(1+z_1)+K_1]}\times\left[\frac{3}{2x_1^3}\left(\sin^{-1}x_1-x_1\sqrt{1-x_1^2}\right)\right]e^{-\frac{(M-\overline{M})^2}{2\sigma^2}}dM\ , \tag{44}$$

where

$$\overline{M}=(M_0-1.3818\sigma^2)\ . \tag{45}$$

This gives us an expression for the number of nebulae having an absolute magnitude between $M$ and $M+dM$ and an apparent magnitude less than $m_1$, the quantities $z_1$, $K_1$, and $x_1$ being the red-shift, correction term, and value of $\bar{r}/R_0$ for a nebula of apparent magnitude $m_1$ and absolute magnitude $M$.

We may now integrate expression (44) over all possible values of $M$ from minus to plus infinity, holding $m_1$ constant and letting $z_1$, $K_1$, and $x_1$ assume the values which they have for this value of $m$ and the varying values of $M$. We thus obtain for the total number of nebulae out to the apparent magnitude $m_1$

$$N_1=\frac{\text{const.}}{\sqrt{2\pi\sigma}}\,10^{0.6m_1}\int_{-\infty}^{+\infty}10^{-0.6[5\log(1+z_1)+K_1]}\times\left[\frac{3}{2x_1^3}\left(\sin^{-1}x_1-x_1\sqrt{1-x_1^2}\right)\right]e^{-\frac{(M-\overline{M})^2}{2\sigma^2}}dM\ , \tag{46}$$

where the constant is independent of the particular value of $m_1$ under consideration.

The only practical method of performing the integration indicated in equation (46) is to use arithmetical or graphical quadrature. To carry this out we can prepare a table with the red-shift $z=\delta\lambda/\lambda$ taken as a convenient choice for the independent variable.

With the help of equation (10),

$$z=\frac{\delta\lambda}{\lambda}=kR_0\sin^{-1}\frac{\bar{r}}{R_0}=kR_0\sin^{-1}x\ , \tag{47}$$

we can first compute the values of $\bar{r}$ and $x$ which correspond to the listed values of $z$. In doing this we take for $k$ the relatively well-determined observational value of Hubble and Humason; on the other hand, in our present state of considerable ignorance as to the density of matter in the universe, we can best regard $R_0$ as an adjustable parameter, and test a variety of choices, including $R_0 = \infty$, which would correspond to no appreciable effects of curvature on nebular counts.

With the help of the methods of section 4*d*, we can next obtain values for the correction term $K$ as a function of the listed values of $z$ which may be represented by

$$z \longleftrightarrow K . \tag{48}$$

In doing this we must take cognizance of the fact that the isolated, non-cluster nebulae to which our counts will apply have the average character of intermediate-type spirals. From a knowledge of the spectra of such objects we can then make an appropriate choice of a reasonable black-body temperature, or spectral class, or combination of spectral classes, by which to represent the unshifted distribution of radiation.

Finally, with the help of equation (21),

$$\log \bar{r} = 0.2\{m - 5 \log (1+z) - K - M\} + 1 , \tag{49}$$

we can introduce a column giving the values of the absolute magnitude $M$ that correspond to the values of $z$, $\bar{r}$, $x$, and $K$ already listed, and to the particular choice of $m = m_1$ which we are studying.

In this way we can determine all the quantities which occur in the integral in (46) as functions of $M$ and thus make the desired quadrature.

By the application of such methods of computation we have found that on account of the fairly sharp maximum exhibited by the exponential factor $e^{-\frac{(M-\overline{M})^2}{2\sigma^2}}$, the integral in (46) can often be satisfactorily approximated by assigning to the rest of the integrand the value that it has at the maximum point where $M = \overline{M}$. We then obtain

$$N_1 = \text{const. } 10^{0.6[m_1 - 5 \log (1+z_1) - K_1]} \left[ \frac{3}{2x_1^3} \left( \sin^{-1} x_1 - x_1 \sqrt{1 - x_1^2} \right) \right] , \tag{50}$$

or, by taking logarithms,

$$\left.\begin{aligned}\log N_1=0.6\Big\{ & m_1-5\log(1+z_1)-K_1\\ & +\tfrac{10}{6}\log\left[\frac{3}{2x_1^3}\left(\sin^{-1}x_1-x_1\sqrt{1-x_1^2}\right)\right]\Big\}+\text{const.}\,,\end{aligned}\right\}\quad(51)$$

where $z_1$, $K_1$, and $x_1=\bar{r}_1/R_0$ are to be given the values which they would have for a nebula of apparent magnitude $m_1$ and absolute magnitude $M=\overline{M}$.

In accordance with (45), the value $\overline{M}$ is related to the mean value $M_0$ for the absolute magnitudes of the nebulae by the equation

$$\overline{M}=M_0-1.3818\sigma^2\,. \qquad (52)$$

Using the actual observational value for the dispersion $\sigma$, we find $\overline{M}$ to be about 1 mag. brighter than the true mean $M_0$. This consequence of the presence of dispersion may be regarded as a selection effect, the absolutely brighter nebulae being favored in the counts as compared with the fainter ones, since they reach to greater distances before their apparent magnitudes assume the particular limit $m_1$ set for the count in question.[7, 10]

Since all the quantities in (51) are empirically determinable, at least as soon as we have chosen a value for the parameter $R_0$, this equation provides a second possibility for observational tests of the model. It is to be noted, however, in contradistinction to the previous possibility provided by equations (39) for the distribution of surface intensity over the images of similar nebulae, that equation (51) involves the assumption of uniform nebular distribution fully as much as that of nebular recession.

## PART II. TREATMENT ASSUMING RED-SHIFT NOT DUE TO RECESSION

### 7. THE THEORETICAL FORMULAE IN THE ABSENCE OF RECESSION

We now turn to a determination of the results to be expected in case the red-shift is due to some cause other than recession. For this purpose it will be sufficient to indicate the modifications in our pre-

[10] See Hubble, *Halley Lecture*, Oxford, 1934.

vious equations arising from the changed point of view. The numbers of the new equations are those of the analogous preceding ones with an accent added.

As already mentioned, the original Einstein static universe is taken as a suitable model for the case now considered in which the red-shift is not due to recession. For the line element we then have

$$ds^2 = -\frac{d\bar{r}^2}{1-\bar{r}^2/R_0^2} - \bar{r}^2 d\theta^2 - \bar{r}^2 \sin^2\theta d\phi^2 + dt^2 . \qquad (1')$$

Since there would be no systematic motion of the nebulae in such a model, we must assume some other unknown cause for the red-shift. In the absence of knowledge as to the mechanism of this cause, we shall take the fractional red-shift as proportional to the proper distance to the nebula, namely,

$$\frac{\delta\lambda}{\lambda} = k\int_0^{\bar{r}} \frac{d\bar{r}}{(1-\bar{r}^2/R_0^2)^{\frac{1}{2}}} . \qquad (2')$$

Somewhat different laws for the red-shift might perhaps seem more plausible, but (2′) provides us with an expression which lies within the limits prescribed by the known data and leads to the same form of relation between red-shift and co-ordinate position as was used in treating the recessional theory.

We assume that the unknown cause for the red-shift acts so as to decrease the energy associated with each photon but not to deflect it appreciably from its geodesic path. For the angular diameters of similar nebulae, located at different values of $\bar{r}$ but simultaneously observed at the origin, we then have

$$\delta\theta = \frac{\text{const.}}{\bar{r}} . \qquad (3')$$

Furthermore, since the new cause for the red-shift is assumed to decrease the energy of photons without affecting their rate of arrival, we may write for the bolometric luminosity of similar nebulae at different co-ordinate positions

$$l_b = \frac{\text{const.}}{\bar{r}^2}\left(\frac{\lambda}{\lambda+\delta\lambda}\right) . \qquad (4')$$

Finally, for the number of nebulae in a given co-ordinate range, we again have

$$dN = \text{const.}\ \frac{\bar{r}^2 d\bar{r}}{(1-\bar{r}^2/R_0^2)^{\frac{1}{2}}}\ . \tag{5'}$$

8. RE-EXPRESSION OF FORMULAE FOR USE WITH OBSERVATIONS IN THE ABSENCE OF RECESSION

The consequences of the foregoing theoretical formulae may now be expressed in forms adapted to the discussion of observational findings, as was done in section 4 for the case of recession. Since the derivations are similar to the earlier ones, we merely state the new formulae which are of special interest.

For the red-shift as a function of co-ordinate position, as before,

$$z = \frac{\delta\lambda}{\lambda} = kR_0 \sin^{-1}\frac{\bar{r}}{R_0} = kR_0 \sin^{-1} x\ . \tag{10'}$$

For the ratio of the observed radii $\rho'$ and $\rho$ for corresponding points on similar nebulae, we have the new expression

$$\frac{\rho'}{\rho} = \frac{\bar{r}}{\bar{r}'}\ . \tag{11'}$$

The new relation between co-ordinate position $\bar{r}$ and photographic magnitude is

$$\log \bar{r} = 0.2\{m - 2.5 \log (1+z) - K - M\} + 1\ , \tag{21'}$$

where $K$ has the same value as before, but the total correction to the photographic magnitude $m$ is smaller than before by the amount $2.5 \log (1+z)$, owing to the fact that there is no longer any recessional action to decrease the rate of arrival of photons.

Finally, for the number of nebulae in the co-ordinate range $d\bar{r}$ having an absolute magnitude in the range $dM$, we again have the expression

$$dN = \frac{C}{\sqrt{2\pi}\sigma} e^{-\frac{(M-M_0)^2}{2\sigma^2}} \frac{\bar{r}^2 d\bar{r}}{(1-\bar{r}^2/R_0^2)^{\frac{1}{2}}} dM\ , \tag{29'}$$

where $C$ is a constant.

## 9. DISTRIBUTION OF LUMINOSITY OVER THE SURFACE OF NEBULAR IMAGES IN THE ABSENCE OF RECESSION

We now consider the relations between nebular luminosities and dimensions to be expected in case the red-shift is not due to recession. Combining (3′) and (4′), we obtain

$$\frac{\delta\theta}{l_b^{\frac{1}{2}}}=\text{const.}\left(\frac{\lambda+\delta\lambda}{\lambda}\right)^{\frac{1}{2}}, \tag{30′}$$

as the relation between observed diameters, bolometric luminosities, and red-shift for similar nebulae at different distances. It differs from the previous expression by the entrance of the red-shift term to the power $\frac{1}{2}$ instead of 2. The result in this simple form, however, applies as before to objects which are thought of as having a sharp cut-off in surface brightness at some definite radius.

For actual elliptical nebulae with a surface brightness which decreases gradually from the center outward, we obtain

$$\frac{i}{i'}=\frac{(1+z')\,p}{(1+z)\,p'}\,. \tag{34′}$$

This expression for the ratio of the intensities of photographic illumination at corresponding points on the images of two essentially similar nebulae at different distances from the observer contains the red-shift terms to the power 1 instead of 4 as in the previous result.

Furthermore, introducing the known observational facts as to the distribution of the surface brightness of elliptical nebulae, we again obtain

$$i=\frac{i_0}{\left(\frac{\rho}{a}+1\right)^2} \tag{37′}$$

for the photographic intensity $i$ as a function of radial position $\rho$, but the parameters $i_0$ and $a$ for similar nebulae at different distances now satisfy the appreciably altered relations

$$\left.\begin{aligned}\log i_0&=\text{const.}-\log(1+z)-0.4K\,,\\ \log a&=\text{const.}-0.2m+0.5\log(1+z)+0.2K\,.\end{aligned}\right\}\,. \tag{39′}$$

Since all the quantities in (37′) and (39′) are empirically determinable, comparison with the results from (37) and (39) should give a possibility of distinguishing between the two types of explanation for the red-shift.

### 10. NEBULAR COUNTS TO A GIVEN LIMITING MAGNITUDE IN THE ABSENCE OF RECESSION

The two types of explanation for the red-shift lead also to different expectations as to the dependence of nebular counts on apparent magnitude. In the absence of recession, we should have

$$\left.\log N_1 = 0.6\left\{ m_1 - 2.5\log(1+z_1) - K_1 + \tfrac{10}{6}\log\left[\frac{3}{2x_1^3}\left(\sin^{-1}x_1 - x_1\sqrt{1-x_1^2}\right)\right]\right\}\right\} \qquad (51')$$

for the number of nebulae $N_1$ brighter than apparent photographic magnitude $m_1$, where $z_1$, $K_1$, and $x_1 = \bar{r}_1/R_0$ are again to be given the values which they would have for a nebula of apparent magnitude $m_1$ and absolute magnitude $\overline{M} = M_0 - 1.3818\sigma^2$. The expression differs from the previous one by containing the term $2.5 \log(1+z_1)$ instead of $5 \log(1+z_1)$, thus providing another possibility of distinguishing between the two types of explanation for the red-shift.

## Part III. Present Status of Observational Work

### 11. INVESTIGATION OF THE STRUCTURE OF NEBULAR IMAGES

Attention has been given in the Mount Wilson program of observation to both the foregoing methods of attack, and a preliminary statement as to the present status of the observations will not be out of place in view of our natural desire to understand the causes for such a general phenomenon as the nebular red-shift.

Consider first the method provided by the theoretical relations for the distribution of intensity over the surfaces of similar nebular images. These relations were derived for the case of elliptical nebulae in which differences between the spectral distribution of light from the center and that from the periphery of the object could be neglected. Since the great nebular clusters are composed largely of elliptical nebulae, all in any one cluster being at approximately the same distance from us, the photographic study of clusters should provide especially suitable data. The investigation can be made by exposing to clusters at different distances for the same length of time and then using the empirical relations between intensity of photographic illumination and resulting photographic effect for a given exposure time.[9]

On the basis of both theories we should find nebular images in the different clusters with photographic intensity $i$ given as a function of radius $\rho$ by relations of the form

$$i=\frac{i_0}{\left(\frac{\rho}{a}+1\right)^2} . \qquad (37, 37')$$

The two theories would lead to different expectations, however, as to the dependence on distance of the parameters $i_0$ and $a$, for a given kind of nebula, as we go to photographs of more and more distant clusters.

We may first inquire whether this difference between the two theories could be detected by measuring the over-all dimensions to which the nebular images in the different clusters build up in their prescribed period of exposure. Denoting by $i_{th}$ the threshold intensity which is just sufficient to give a detectable effect in the prescribed period, we can solve (37, 37′) for the peripheral radius $\rho_{th}$ in the form

$$\rho_{th}=a\left(\frac{i_0^{\frac{1}{2}}}{i_{th}^{\frac{1}{2}}}-1\right) . \qquad (53)$$

And since the central intensity $i_0$ will be very large compared with the threshold intensity $i_{th}$, (53) can be written in the form

$$\log \rho_{th}=\log a+0.5 \log i_0-0.5 \log i_{th} . \qquad (54)$$

For different distances, however, the threshold intensity $i_{th}$ will be the same, while for $a$ and $i_0$ we can substitute expressions (39) and (39′), which relate these parameters to distance on the basis of the two theories. We thus obtain for both theories exactly the same form of expression

$$\log \rho_{th}=\text{const.}-0.2m . \qquad (55)$$

In accordance with this expression the over-all dimensions of the images would decrease according to the same law for both theories as we go to fainter and hence larger magnitudes. This necessary conclusion is unfortunate since a determination of the over-all dimensions of nebular images with the help of microphotometric analysis is the most practical measurement of their structure that we can make.

The preliminary observations on cluster nebulae thus far made at Mount Wilson have been in agreement with equation (55). Even within any given cluster the dimensions and magnitudes of the nebulae are found to be connected by an approximately linear relation of the form (55), presumably on account of some dependence of total luminosity on dimensions; and then on passing from one cluster to a more distant one the same relation continues to hold without appreciable change in the value of the constant. This check of a relation which would be expected to hold on both theories can serve to give an increased feeling of confidence in our general picture of nebulae and nebular clusters as really being extra-galactic objects at different distances from us, but cannot distinguish between the two suggested explanations of the red-shift.

To obtain criteria for distinguishing the two types of explanation, with the help of our analysis of the surface distribution of nebular luminosity, a more detailed study of the parameters $i_0$ and $a$ for the nebulae in different clusters would be necessary. Expressions (39) and (39′) for these quantities in the case of essentially similar nebulae at different distances can be re-written in the forms

$$\left.\begin{aligned} i_0(1+z)^4 10^{0.4K} &= \text{const.} \\ \frac{a 10^{0.2m}}{(1+z)^2 10^{0.2K}} &= \text{const.} \end{aligned}\right\} \text{(with recession)}$$

and

$$\left.\begin{aligned} i_0(1+z) 10^{0.4K} &= \text{const.} \\ \frac{a 10^{0.2m}}{(1+z)^{\frac{1}{2}} 10^{0.2K}} &= \text{const.} \end{aligned}\right\} \text{(without recession)}$$

Since all the quantities in these formulae are measurable, the quite different dependence on the red-shift $z = \delta\lambda/\lambda$ for the two theories should be detectable, at least in principle, by determining the distribution of the foregoing quantities among the individual nebulae of clusters at different distances.

Unfortunately, this program of possible investigation is made very difficult by the necessity of going to quite distant clusters in order to get sufficiently large red-shifts. The nebular images then become so small compared with microphotometer aperture and size of plate grain that determinations of the central intensity $i_0$ or of the radius $a$ where the intensity has fallen to one-quarter are not easy. It is to be hoped, nevertheless, that further development of technique with the 100-inch telescope or the use of the 200-inch telescope now under construction will at some time make this method of study applicable.

## 12. DETERMINATION OF NEBULAR COUNTS

We now turn to the method of investigating the nature of the red-shift provided by the determination of nebular counts to successive limits of apparent magnitudes. Here, in accordance with equations (51) and (51′), the two theories lead to an expression for the number of nebulae $N_1$ to a given limiting magnitude $m_1$ of the same form,

$$\log N_1 = 0.6(m_1 - \Delta m_1) + \text{const.}, \tag{56}$$

where, however, the correction $\Delta m_1$ to be applied to the apparent magnitude $m_1$ depends on the type of explanation proposed for the red-shift.

In the absence of any red-shift and with a uniform distribution of nebulae in Euclidean space, the correction $\Delta m_1$ would, of course, be zero, and the foregoing expression would reduce to the familiar form

$$\log N_1 = 0.6 m_1 + \text{const.} \tag{57}$$

The fact that this simple equation does give an approximate expression for the observational results[11] out to distances where the effects of red-shift and presumable spatial curvature would be unimportant may be taken as agreeing with our general picture of the extra-galactic distribution of nebulae.

Allowing, however, for the actual presence of a red-shift in the

[11] Hubble, *Mt. W. Contr.*, No. 485; *Ap. J.*, **79**, 8, 1934.

light from the nebulae and treating them as uniformly distributed in a space of constant curvature, we have found the correction term to have the form

$$\Delta m_1 = 5 \log (1+z_1) + K_1 - \tfrac{10}{6} \log \left[ \frac{3}{2x_1^3} \left( \sin^{-1} x_1 - x_1 \sqrt{1-x_1^2} \right) \right] \quad (58)$$

for the case of recession as the cause of the red-shift, and the form

$$\Delta m_1 = 2.5 \log (1+z_1) + K_1 - \tfrac{10}{6} \log \left[ \frac{3}{2x_1^3} \left( \sin^{-1} x_1 - x_1 \sqrt{1-x_1^2} \right) \right] \quad (58')$$

on the basis of a non-recessional theory of the phenomenon. In both of these expressions the fractional red-shift $z_1$, the correction term $K_1$, and the ratio $x_1 = \bar{r}_1 / R_0$ are to be taken in accordance with our previous discussion as the values for a nebula having the apparent magnitude $m_1$ and the absolute magnitude

$$\overline{M} = M_0 - 1.3818\sigma^2 , \quad (59)$$

where $M_0$ is the mean absolute magnitude for isolated nebulae and $\sigma$ the dispersion around this mean.

In order to make tests of the differences implied by equations (58) and (58′), it is necessary to have a homogeneous body of nebular counts to a series of different limiting magnitudes, which go out far enough to make the effects of red-shift appreciable. For each limiting magnitude the areas of the sky covered must be so selected as to give a representative sample of isolated non-cluster nebulae, and the different limiting magnitudes themselves must be determined by methods having sufficient accuracy and interconsistency. The attainment of such data forms a definite part of the Mount Wilson program of observation.

At present the nearest approach to a satisfactory body of data is provided by published[12] and unpublished[13] nebular surveys made at Mount Wilson with several different exposure times at the 60-inch

[12] Hubble, *Mt. W. Contr.*, No. 485; *Ap. J.*, **79**, 8, 1934, giving surveys to $m=19.4$ and $m=20.0$, made with one-hour exposures at the 60-inch and 100-inch reflectors.

[13] Surveys by Hubble to $m=18.5$ and $m=21.0$ made with twenty-minute exposures at the 60-inch and two-hour exposures at the 100-inch reflector.

and 100-inch reflectors, combined with the survey of Mayall[14] at Mount Hamilton based on one-hour exposures with the 36-inch reflector. Each of these surveys provides a set of plates, on which the numbers of nebulae have been counted down to a definite threshold size and density of image, so chosen that the distinction between stars and nebulae is still sufficiently reliable, together with an independent determination of the apparent magnitudes of the nebulae giving these threshold images. In this way values have been obtained for the average number of isolated nebulae per square degree brighter than a series of limiting magnitudes estimated as $m=18.5$, 19.0, 19.4, 20.0, and 21.0.

The time and effort needed for obtaining such a set of surveys, together with the necessity for unique telescopic facilities which must also be used for other tasks as well, make it desirable to employ the present data to the fullest possible extent. This makes it necessary to draw tentative conclusions which are affected by difficulties and uncertainties as to the best analysis of the present data and may require serious modification as further data accumulate. We must now consider the nature of some of these difficulties and uncertainties. We shall take up the different quantities involved in the order in which they occur in the equations to be tested.

Turning first to the nebular counts $N_1$, it will of course be obvious that the actual process of determining the number of nebular images per plate is difficult and laborious, involving the use of systematic methods of counting and checking, as well as the maintenance throughout each survey of a uniform threshold of identification at a level giving at least an approximate balance between the numbers of stellar and nebular images wrongly identified. In addition, however, in order to secure a homogeneous body of data, the actual count on each plate must be subjected to a number of corrections to make it correspond to some selected set of standard conditions. In practice[11] this means multiplication by a series of factors to obtain a corrected count corresponding to the atmospheric extinction at the zenith, to the galactic obscuration at the galactic pole, to plates of "excellent" quality, to the standard time of exposure selected for the

[14] *Lick Obs. Bull.*, **16**, 177 (No. 458), 1934, giving a survey to $m=19.0$, made with one-hour exposures at the 36-inch reflector.

survey, and to the photographic definition at the center of the plate. Some uncertainty is undoubtedly introduced by the necessity for such elaborate correction; moreover, this uncertainty must not be minimized since the corrections up to "excellent" quality of plate are so much more important for the most distant as compared with the nearest survey that quite different conclusions would be drawn from a direct use of the uncorrected counts. As a final remark concerning $N_1$, it may be mentioned that the different plates of a survey, especially of surveys for the less distant nebulae, show a wide spread in the corrected counts. Nevertheless, a study of the character of the dispersion shows that the actual mean counts give values in good agreement with what would be expected for an infinite collection of such survey plates.

Turning next to the determination of the limiting magnitudes $m_1$ for the different surveys, we meet the usual difficulties in accurate magnitude determinations, and, in addition, a special difficulty connected with the very faint magnitudes to which the surveys must be carried in order to make the desired tests. The most reliable determinations of magnitudes are to be obtained with the help of schraffierkassette images. And for the nearer surveys it is possible with sufficient exposure time to obtain such images for objects that do give threshold images under survey conditions. The more distant objects, however, are too faint to give such images with exposure times that are possible, and methods of extrapolation must be used. These methods, which cannot be discussed here, have, to be sure, led to results of considerable internal consistency. Nevertheless, an error thus introduced which was itself a function of magnitude might be serious.

We must next consider the fractional red-shift $z = \delta\lambda/\lambda$, which is a quantity of particular interest because of its different occurrence in the alternative expressions for the correction to photographic magnitude, $\Delta m_1$, given by the two theories. With the help of spectroscopic and photometric measurements made on selected clusters, the relation between red-shift and magnitude has now been determined for red-shifts up to about 0.13. The results show a linear dependence of $\delta\lambda/\lambda$ on the co-ordinate $\bar{r}$, within limits of accuracy such that either of the formulae, (21) or (21′), given by the two theories may be employed for the calculation of the co-ordinate $\bar{r}$ from the measured magnitude $m$. Hence, with allowances for the possibility that spatial curvature may have an effect at still larger distances, it seems reasonable, as already suggested, to make the extrapolation to greater distances with the help of equation (10). Nevertheless, the uncertainties involved in the long extrapolation must not be minimized. For the nearest survey, to be sure, the value of $\delta\lambda/\lambda$ is found to be only about 0.08 for a nebula having the apparent magnitude $m = 18.5$ and the absolute magnitude $M = \overline{M}$, which in accordance with our discussion of equation (46) is a suitable "average" nebula to use in calculating the correction term $\Delta m_1$. On the other hand, for the most distant survey the value of $\delta\lambda/\lambda$ would be about 0.20 for a nebula having the apparent magnitude $m = 21.0$ and the absolute magnitude $M = \overline{M}$, which lies well beyond the limit of about 0.13 up to which measurements have been made. Furthermore, it is to be emphasized that the precise evaluation of the integral occurring in equation (46) is appreciably dependent on the value of $\delta\lambda/\lambda$ for a range of distances on both sides of that corresponding to $m_1$ and $\overline{M}$. Hence the necessity of extrapolation introduces uncertainties which are unavoidable at present but may be serious.

We must now turn to difficulties connected with the correction term $K_1$. The methods for relating this quantity to an assumed spectral type or an assumed black-body temperature for the nebulae involved have already been discussed in section 4*d*. Difficulties arise, however, as to what spectral type or black-body temperature to choose. The isolated nebulae involved in the counts range through all types with a reasonable average around Sb spirals. The greater proportion of the effective radiation from such spirals is known to come from the inner portions of the nebulae and to have a spectrum similar to that from stars of approximately the type dG2, which corresponds to an effective temperature of about $T_0 = 5840°$. There are two complicating factors, nevertheless, which affect the situation in opposite directions. First, there are blue stars of much earlier type in the outer regions of such nebulae which would raise the effective temperature by an unknown but perhaps not very large amount; while, on the other hand, unpublished measures with photoelectric cells and color filters made by Stebbins and his associates at the foci of the two Mount Wilson reflectors have given color indices for the inner regions of such nebulae which would correspond to effective temperatures perhaps as low as $T_0=4860$. Further work, including the use of red-sensitive plates and color screens for the photographic determination of the color indices of distant nebulae, may clarify the situation. In the meantime, however, any statement as to the conclusions to be drawn from nebular counts have to be judged in the light of some accompanying statement as to the spectral type or effective temperature assumed in the calculation of the correction term $K_1$.

We turn now to the final constant term occurring in equation (56) for the expected number of nebulae to different limiting magnitudes. Here it may be remarked that the constant could be determined, either to give the best fit for the five surveys under consideration or in such a way as to give weight to counts that have been made to numerically smaller magnitudes at other observatories. The assignment of proper weight to such other results is difficult, however, owing to differences in methods of magnitude determination and to the possibility that the number of nebulae in our immediate neighborhood may be affected by a fluctuation from the mean.

To complete the discussion of the complexities which affect our present method of testing the red-shift, we must also mention the mean absolute magnitude $M_0$ and the dispersion $\sigma$ which determine, in accordance with equation (59), the absolute magnitude $\overline{M}$ of a suitable "average" nebula to use in calculating the corrections $\Delta m_1$ to the photographic magnitudes. These two quantities, $M_0$ and $\sigma$, must themselves be determined from an analysis of rather limited data on the absolute magnitudes of nearby isolated nebulae, and there is some uncertainty as to how well the distribution can be represented by a Gaussian curve and how accurate the calculated values of the mean and the dispersion may be.

The foregoing long discussion of the difficulties and uncertainties affecting the treatment of nebular counts has seemed necessary in order to emphasize the provisional character of any conclusions as to the nature of the red-shift which can be drawn from such counts at the present time. The tentative conclusions indicated by an analysis of the five surveys to different magnitudes mentioned above may be stated quite briefly.

In the first place, it is quite certain that the counts cannot be represented by the simple formula (57) which corresponds to stationary objects in Euclidean space without red-shift at all, and that to attain agreement some correction term, $\Delta m_1$, which increases with distance, must be subtracted from the limiting magnitudes $m_1$ as in equation (56). Hence, at least we can say that the red-shift does appear to cut down the luminosity of distant nebulae in the general way that might seem theoretically inevitable. In the second place, nevertheless, it is not so certain that the values of the corrections $\Delta m_1$ actually needed to secure agreement with the data are large enough to be regarded as theoretically plausible.[10]

In the case of the non-recessional theory, which predicts smaller effects from the red-shift, approximate agreement with the data could be attained, provided we disregard the possibility of any appreciable light-absorption on the way from the nebulae, and then take for the correction $\Delta m_1$ values calculated from equation (58′) with the aid of a $K_1$ corresponding to the high effective temperature of about 6000°, and at the same time neglect any possible effects of spatial curvature. On the other hand, in the case of the recessional theory, which predicts larger effects from the red-shift, approximate agreement with the data would seem attainable, in accordance with equation (58), only by also making use of possible effects from spatial curvature which correspond to the appearance of the third term in the expression for the correction $\Delta m_1$.

In either case the necessity of neglecting any effect of absorption over the long light-path from the nebulae and of getting low values for $K_1$ by accepting effective temperatures for the nebulae which seem high in view of the present observations on color indices must be regarded with some uneasiness. In addition, in the case of the recessional theory, the necessity for also making use of spatial curvature to cut down the calculated values of the correction $\Delta m_1$ must be regarded as in conflict with our usual notions as to the distances to which observations would have to be carried before appreciable effects from spatial curvature would seem probable.

As to the nature of the spatial curvature, which would be needed in the case of the recessional theory, it is to be noted—as shown by the form of the original equation (5) for the number of nebulae in a given co-ordinate range—that the effects of curvature would be in the right direction to give increased counts at larger distances only for the case of a closed model with $R_0$ real. This result is of interest in connection with Milne's work, since he has shown[15] that his kinematical model would lead to nebular counts approximating those in a relativistic model with negative curvature, that is, with $R_0$ imaginary instead of real, as appears observationally indicated. As to the amount of curvature, it is found with the help of the usual relativistic expressions for pressure and density in a homogeneous cosmological model that the necessary value of $R_0$ would require an averaged-out density in the universe several thousand times as large as that computed from the masses usually assigned to the nebulae. Such a density would necessitate either large amounts of unknown intergalactic material in a form which would not lead to appreciable absorption, or very large masses connected with the nebulae themselves, as is perhaps suggested by the recent work of Sinclair Smith[16] on the mass of the Virgo cluster.

Our conclusions as to the present status of nebular counts can therefore be summarized by stating that it might be possible to explain the results on the basis of either a static homogeneous model with some unknown cause for the red-shift or an expanding homogeneous model with the introduction of effects from spatial curvature which seem unexpectedly large but may not be impossible. This statement must be made very tentatively, however, and may need serious modification in the light of further counts, of new auxiliary data for treating the counts, or of better analysis of the present data.

## 13. CONCLUDING REMARKS

To bring this long treatment to an end a few further remarks are needed. It is to be noted that both types of explanation for the nebu-

[15] *Relativity, Gravitation and World Structure* (Oxford, 1935), p. 333, eq. (ii‴).

[16] *Pub. A.S.P.*, **47**, 220, 1935.

lar red-shift which we have examined have been treated with the help of models that were assumed to be filled with a uniform distribution of the different kinds of nebulae, having properties sufficiently independent of time to permit the nebulae at all distances being regarded as similar at the time their observed light was emitted.

With regard to the assumed uniformity of nebular distribution, the theoretical possibility for testing the nature of the red-shift by measurements on the surface distribution of nebular luminosity may again be stressed, since the results of such a test should be less dependent on the assumed homogeneity of the model than those derived from nebular counts. As to the possibility of assuming actual deviations from uniform distribution to explain the unexpectedly large nebular counts at great distances, nothing entirely convincing can be advanced against such an assumption. Nevertheless, the analysis already published for the very extensive nebular counts to $m=19.4$ and $m=20.0$ shows no evidence for lack of uniformity in different directions. Hence, to make use of non-homogeneity as an explanation in the present connection, we might have to regard ourselves as rather uniquely located near the center of a spherically symmetrical nebular distribution with a density gradually increasing outward, which seems an *ad hoc* hypothesis.

With regard to the assignment of constant properties to the nebulae over long periods of time, it is to be remarked that in the case of the most distant survey, to $m=21.0$, the "average" nebulae involved must have emitted the light which is now observed at a time of the order of $3\times10^8$ years in the past. If the recessional explanation of red-shift were adopted, this period would be so nearly comparable with the time of cosmic expansion—possibly of the order of $10^9$ to $10^{10}$ years—that the assignment of constant properties might not appear justified, and in that case we might try to explain excess counts by assuming greater luminosities for the nebulae at earlier times. On the other hand, with the non-recessional theory, there would be nothing to indicate a short-time scale for nebular evolution, which indeed might seem to be an advantage for that theory.

In conclusion it may be repeated that the main function of the present article has been to provide a formulation for the two theoretical methods of investigating the nature of the red-shift which would be applicable under actual circumstances, the length of the article being due to the real complexity of these circumstances. In addition it seemed of interest to give some tentative statement as to the present status of the data for the tests, even though their complete presentation and analysis must be reserved for a later time. It also seemed desirable to express an open-minded position as to the true cause of the nebular red-shift, and to point out the indications that spatial curvature may have to play a part in the explanation of existing nebular data.

Carnegie Institution of Washington
Mount Wilson Observatory,
California Institute of Technology
July 1935