PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

Do We Really Need Dark Matter? [and Discussion]

J. J. Binney and R. J. Tayler

Phil. Trans. R. Soc. Lond. A 1986 **320**, 465-473
doi: 10.1098/rsta.1986.0129

Phil. Trans. R. Soc. Lond. A **320**, 465–473 (1986)
Printed in Great Britain

Do we really need dark matter?

By J. J. Binney

Department of Theoretical Physics, Keble Road, Oxford OX1 3NP, U.K.

Stars that between them contain a very small proportion of the mass in a typical stellar population, generate the bulk of the population's luminosity. Consequently the overall mass:light ratio Υ of the population depends on exactly how much of the population's mass is in luminous stars, and we should not be surprised if stellar populations that formed in different physical conditions were characterized by very different values of Υ. Actually, the well-observed regions of galaxies show minimal variation of Υ. It is argued that even the most extended H I rotation curves could be understood in terms of constant Υ if estimates of the brightness of the night sky were subject to systematic errors in the region of 4–8%.

1. Introduction

Many of science's greatest triumphs derive from the following two-stage process; (i) we seek regularities in natural phenomena and then (ii) predict the results of new measurements by assuming that regularities observed in a restricted range of phenomena apply equally in more general circumstances. Step (ii) leads to striking predictions only if the regularities detected in step (i) are incompatible with many logical possibilities. In other words, important scientific advances emerge from detecting and then extending regularities of low *a priori* probability. I shall argue that the mass:light ratios of stellar systems show just such a regularity, and that it may be premature to discount the predictive power of this regularity in favour of alluring speculations about 'dark matter'. The Universe may be a more prosaic place than our high-energy friends would have us believe.

2. Mass and light in stellar populations

A general property of stellar populations is that only a very small part of the total mass of any stellar population is contained in the stars that generate almost all the population's light output. For example, in the solar neighbourhood 98% of the light is emitted by stars that between them contain less than 3.7% of the mass known to be in stars (Allen 1973). The bulk of the luminosity is always emitted by the more massive stars, whereas it is the low-mass stars that contain most of the mass. Hence the mass:light ratio of the population as a whole depends sensitively on the abundance in the population of high- and low-mass stars.

We do not yet understand what factors control the spectrum of stellar masses in a freshly formed population of stars. However, the factors we expect to be important determinants include the (i) density ρ, (ii) metallicity Z and (iii) the pressure p of the gas from which the population formed. Thus we should expect stellar populations that formed from bodies of gas at very different densities, metallicities and pressures would have very different mass:light ratios.

We now have excellent reason to believe that the densities, metallicities and pressures under which stars form in different galaxies do indeed cover enormous ranges. Consider, for example, the contrasting conditions at the centre of a first-ranked cluster galaxy such as NGC 1275 and in the outer regions of an early-type disc galaxy such as the Small Magellanic Cloud (SMC). At the centre of NGC 1275 we have $\rho \approx 7\times10^{-21}$ g cm^{-3}, $Z \approx 1.5Z_\odot$ and $p/k \gtrsim 10^7$ K cm^{-3}, whereas at the de Vaucouleurs radius, R_{25}, in the SMC we have $\rho \lesssim 3\times10^{-24}$ g cm^{-3}, $Z \approx 0.1\ Z_\odot$ and $p/k \lesssim 10^4$ K cm^{-3}. Thus we have a clear *a priori* expectation that the mass:light ratios of galaxies should span an equally extensive range.

3. Mass:light ratios from optical data

Contrary to this expectation, all data concerning the well-observed inner portions of galaxies are compatible with exactly those mass:light ratios that we obtain for these systems by blindly extrapolating from observations of the solar neighbourhood. Three examples will illustrate this proposition:

(*a*) *The rotation curve of our galaxy*

From the work of Oort (1964), Hill *et al.* (1979) and Bahcall (1984*a*, *b*, this symposium) we have that the mass:light ratio of all the material in a column from the Sun to 700 pc† above the galactic plane, is Υ_V (disc) $\approx 5\Upsilon_\odot$. I take this as the mass:light ratio of the disc of any Sb galaxy. Stellar systems such as the spheroid of our Galaxy that lack gas and dust, and have consequently ceased to form significant numbers of young stars, are expected to have larger mass:light ratios; I adopt the value Υ_V (spheroid) $\approx 8.5\ \Upsilon_\odot$, which allows for (i) the dearth in these populations of stars of type earlier than G2 (this raises Υ by about a factor of two), and (ii) their freedom from interstellar dust and gas (which lowers Υ by about 15%). Multiplying estimates of our Galaxy's disc and spheroid luminosity densities from de Vaucouleurs & Pence (1978) by these mass:light ratios, I obtain a mass model of our Galaxy which has the circular-speed curve shown in figure 1 (full line). Also shown in figure 1 (dashed curve) is the empirical rotation curve of Burton & Gordon (1978) (rescaled to a local circular

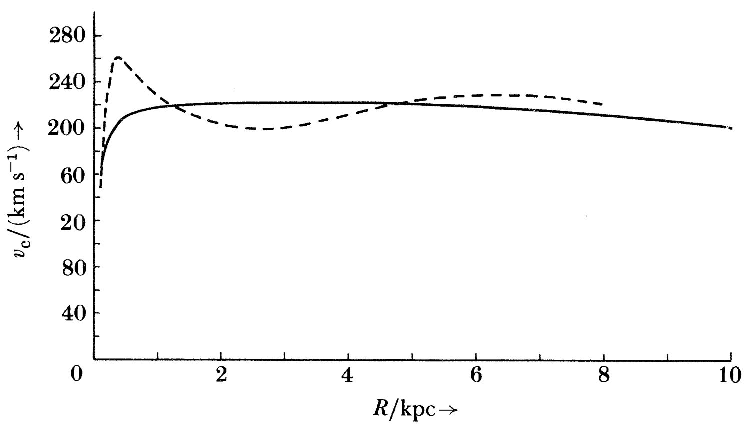

FIGURE 1. The full curve is the galactic rotation curve predicted by the luminosity model of de Vaucouleurs & Pence (1978) if Υ_V (spheroid) = 8.5 $\Upsilon_\odot$ and Υ_V(disc) = 5 $\Upsilon_\odot$. The spheroid and disc luminosities in this model are L_V(spheroid) = $6.3\times10^9\ L_\odot$ and L_V(disc) = $1.1\times10^{10}\ L_\odot$. The dashed curve is the rotation curve derived by Burton & Gordon (1978) from H I and CO line measurements.

† 1 pc ≈ 30857 × 10^{12} m.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A
MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

speed of $v_c = 220$ km s^{-1} (Knapp *et al.* 1978)). Given all the observational uncertainties, and that the fit involves *no free parameter*, the agreement between the theoretical and the empirical curves in figure 1 must be considered remarkable.

(*b*) *The rotation curves of* Sb *and* Sc *galaxies*

As Professor van Albada has shown us, the great majority of the rotation curves of the Sb and Sc galaxies studied by Vera Rubin and her collaborators (Rubin *et al.* 1982), can be accounted for in some detail by assuming that the discs and bulges of these galaxies are characterized by position-independent mass:light ratios comparable to those that fit the rotation curve of our Galaxy.

(*c*) *The kinematics of NGC* 4697

Can the dynamics of elliptical galaxies be reconciled with mass:light ratios similar to those required to account for observations of spiral galaxies such as the Milky Way? R. Davies, G. Illingworth and I have recently modelled three elliptical galaxies, NGC 720, 1052 and 4697, for which CCD R-band photometry and several long-slit spectra are available. The modelling procedure is as follows. (i) We assume the galaxay has an oblate figure of rotation, with some plausible inclination angle i. (ii) We assume some distance D to the galaxy and employ an iterative algorithm based on Lucy's (1974) scheme, to construct from the photometry a three-dimensional luminosity density $j_\mathrm{R}(R, z)$. (iii) The mass density $\rho(R, z)$ is next obtained by multiplying j_R by the mass:light ratio $\varUpsilon_\mathrm{R} = 6.27\varUpsilon_\odot$ that corresponds to $\varUpsilon_\mathrm{V} = 8.5\varUpsilon_\odot$ for a galaxy of colour V−R = 0.85. (iv) From ρ we obtain the gravitational potential $\varPhi(R, z)$. (v) Following Satoh (1980) we integrate the Jeans equations to find the mean-square azimuthal and radial velocities $\overline{v_\phi^2}(R, z)$ and $\overline{v_R^2}(R, z)$ in the system on the assumption that $\overline{v_R^2} = \overline{v_z^2}$. As is well known, this last assumption is equivalent to assuming that the distribution function f of the system is a function $f(E, L_z)$ of the energy and the z-component of angular momentum only. (vi) Because the mass density is strictly independent of how $\overline{v_\phi^2}$ is partitioned into a contribution $\bar{v}_\phi^2$ from mean streaming, and a contribution $\sigma_\phi^2 \equiv (\overline{v_\phi^2} - \bar{v}_\phi^2)$ from random velocities, we set $\bar{v}_\phi = \alpha\surd(\overline{v_\phi^2} - \overline{v_R^2})$, where α is a position-independent parameter, and project $\bar{v}_\phi$ and the appropriate combination of σ_ϕ^2 and $\overline{v_R^2}$ on to the sky for comparison with the measured kinematic quantities. In principle the fit can be adjusted with three free parameters, i, D and α. However, we have independent estimates of D, and the results are probably not very sensitive to the assumed value of the inclination i, so we have so far confined ourselves to a single arbitrarily chosen value of i.

The sort of results that are obtained in this way are illustrated by figure 2 (*a*, *b*), which shows model and observed velocities for NGC 4697 in the case $i = 80°$, $D = 14$ Mpc and $\alpha = 0.75$. The distance $D = 14$ Mpc would be in perfect agreement with the redshift $cz = 1249$ km s^{-1} if $h = 0.89$ and no correction were required for the Virgocentric flow. Clearly this is as plausible a distance as any other, and yet it will be seen that the theoretical curves in figure 2 intersect the great majority of the 2σ error bars on the observational points. Fits of this quality must be considered to lend support to the hypothesis that $j_\mathrm{V} \propto \rho$ with the constant of proportionality $\varUpsilon_\mathrm{V} \approx 8.5\varUpsilon_\odot$ that we have derived from studies of the solar neighbourhood.

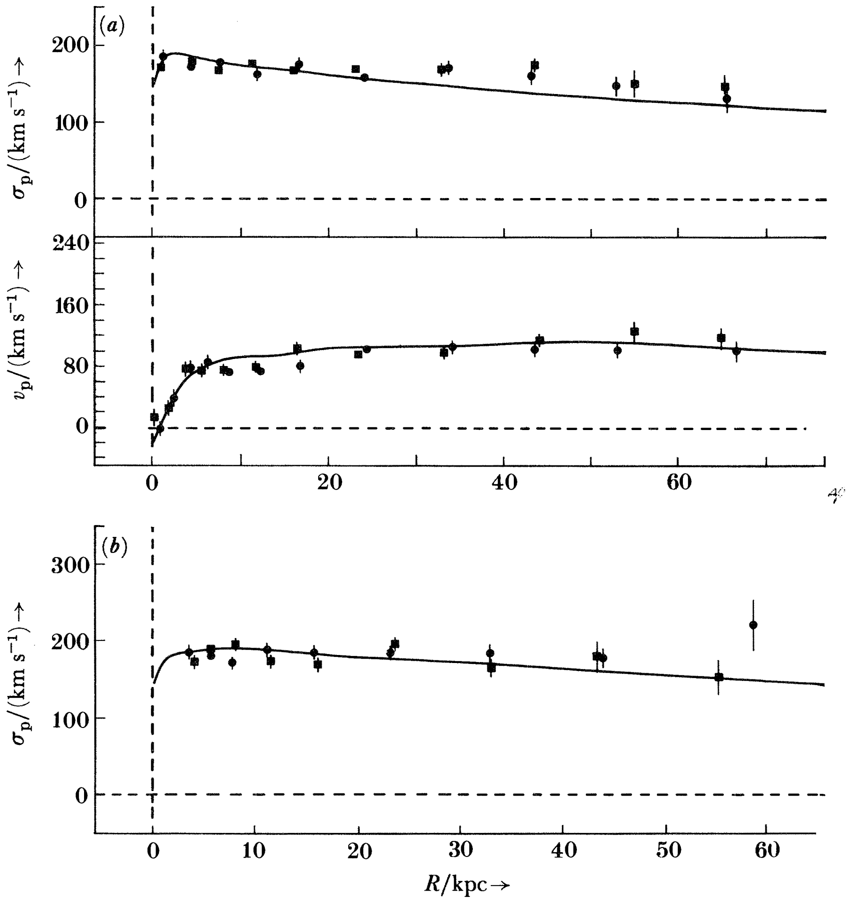

FIGURE 2. Projected velocity dispersion σ_p and streaming velocity $\bar{v}_p$ in the elliptical galaxy NGC 4697 along (*a*) the major axis, and (*b*) the minor axis. (No rotation is detected along the minor axis.) The theoretical curves are for distance $D = 14$ Mpc and mass:light ratio $\Upsilon_R = 6.27\ \Upsilon_\odot$, which corresponds to $\Upsilon_V = 8.5\ \Upsilon_\odot$ for a galaxy of the colour of NGC 4697.

4. The mass density of the disk of NGC 3198

These examples strongly suggest that when a stellar population forms at any position in a galaxy of any Hubble type, it is endowed with high- and low-mass stars in just the proportions characteristic of the young stellar populations of the solar neighbourhood. This is surely a remarkable regularity of low *a priori* probability, which we should seek to extend beyond the domain of its discovery. Let us therefore re-examine the beautiful measurements of H I around the Sc galaxy NGC 3198 that were discussed by van Albada (this symposium) in the light of the question 'are these measurements certainly incompatible with all the mass of NGC 3198 being contained in a standard stellar population?'

The hypothesis we wish to test can be examined by either (i) using photometry to predict the kinematics or (ii) using the kinematics to predict the photometry. In the present case the second course is to be preferred because at large radii the photometry of galaxies like NGC 3198 is a good deal less certain than the kinematic measurements.

NGC 3198 shows little sign of possessing a luminous spheroidal component, so I shall proceed under the assumption that all the light and mass of this system are contained in a flat disc.

The surface density $\Sigma(R)$ at radius R in the disc that generates a given run of circular speed $v_c(R)$ is given by

$$\Sigma(R) = \frac{1}{\pi^2 G}\int_0^\infty H(R,R')\,\mathrm{d}v_c^2(R'), \tag{1}$$

where
$$H(R,R') \equiv \begin{cases} R^{-1}K(R'/R) & \text{if } R' < R \\ R'^{-1}K(R/R') & \text{if } R' > R, \end{cases} \tag{2}$$

and K is the complete elliptic integral of the first kind. Unfortunately this formula does not readily lend itself to the interpretation of observational data, because the logarithmic singularity of $K(k)$ at $k = 1$ causes $\Sigma(R)$ to depend very strongly on the ill-determined gradient of v_c^2 near R. Hence a direct application of (1) to observational data is likely to lead to implausible (and even negative) surface densities $\Sigma(R)$. However, smooth non-negative surface densities $\Sigma(R)$ can be derived from an observed rotation curve $v_c(R)$ as follows.

Let $F_n(R)$ be the radial force at R that is generated by the disc with surface density $\Sigma_n(R)$, and $F_{obs}(R) = v_c^2/R$ be the 'observed' radial force. Then F_n and Σ_n are related by

$$F_n(R) = 2\pi G\int_0^\infty g(R|R')\Sigma_n(R')\,\mathrm{d}R', \tag{3}$$

where
$$g(R|R') \equiv \frac{R'}{\pi R(R+R')}\left[K(k)+\left(\frac{R+R'}{R-R'}\right)E(k)\right]; \quad k \equiv \frac{2\surd(RR')}{R+R'}, \tag{4}$$

and the integral over the singular kernel g is to be interpreted as a Cauchy principal value. By analogy with Lucy's (1974) method, the kernel g, which is not non-negative but does satisfy the normalization condition $\int_0^\infty g(R|R')\,\mathrm{d}R = 1$, may be used to derive from Σ_n an improved surface density Σ_{n+1} according to

$$\Sigma_{n+1}(R') = \Sigma_n(R')\,\exp\left\{\int_0^\infty \ln\left[\frac{F_{obs}(R)}{F_n(R)}\right] g(R|R')\,\mathrm{d}R\right\}. \tag{5}$$

This prescription for generating a series of surface density distributions whose associated forces tend towards the observed forces F_{obs}, has two attractive properties: (i) all iterates are necessarily smooth and non-negative; (ii) if Σ_n differs from the true density Σ_0 by a constant, the normalization of g ensures that $\Sigma_{n+1} = \Sigma_0$. The only snag seems to be that the scheme is unstable at small radii unless care is taken with the extrapolation of Σ_n and the ratio F_{obs}/F_n from the observed domain towards the origin. Stability may be assured by assuming that Σ_n is constant interior to the first grid point at $R = R_{min}$, that $\ln(F_{obs}/F_n)$ is proportional to R for $R < R_{min}$, and by multiplying the argument of the exponential in (5) by a factor that tapers smoothly from unity at $R > R_{min}(R_{max}/R_{min})^{\frac{1}{5}}$ to about 0.1 at $R = R_{min}$.

Figures 3 and 4 show the results of applying this scheme to the rotation curve of NGC 3198 discussed by van Albada (this symposium); the dots in figure 3 show the observed rotation velocities, the lower curve shows the rotation curve of the 'maximum exponential disk' Σ_{em} of van Albada *et al.* (1985), and the upper curve shows the rotation curve generated by the density distribution Σ_{20} that is generated by the scheme described above when $\Sigma_1 = \Sigma_{em}$. The radial scale is set by assuming that the same distance $D = 9.2$ Mpc as van Albada *et al.* The agreement between the upper curve in figure 3 and the data may be considered entirely satisfactory.

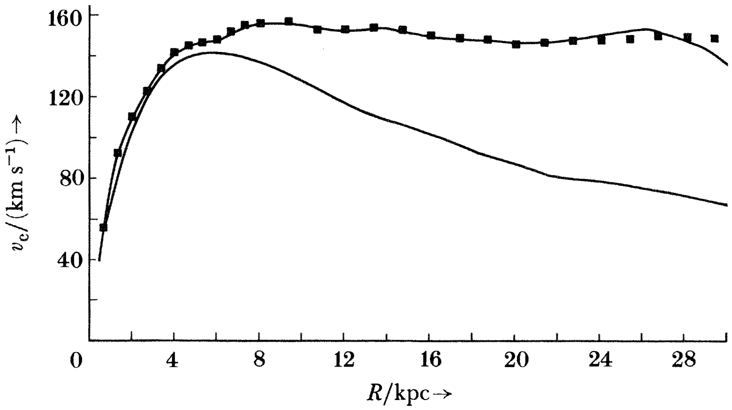

FIGURE 3. The dynamics of the spiral galaxy NGC 3198. The dots show the H I velocities of Begeman (1986). The lower curve is the rotation curve generated by the 'maximum exponential disk' of van Albada *et al.* (1985). The upper curve is the rotation curve obtained as the 20th iterate of the scheme described in the text.

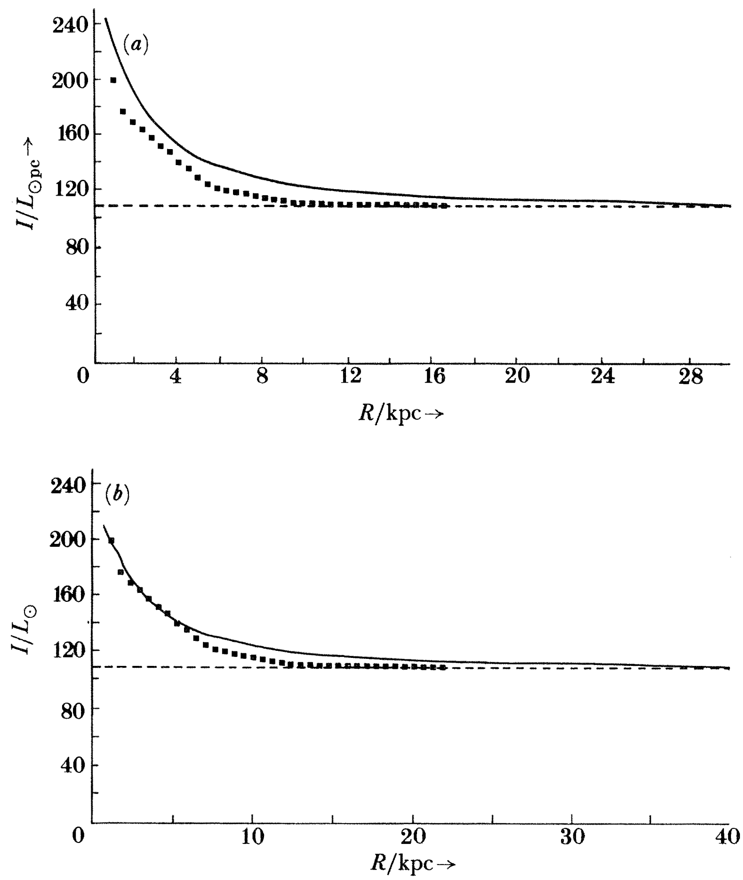

FIGURE 4. (*a*) Comparison of Wevers' (1984) surface brightness measurements corrected to face-on aspect (squares) with the surface brightness predicted by the rotation curve shown in figure 3 if (i) $\Upsilon = 5\ \Upsilon_{\odot}$ and (ii) the distance $D = 9.2$ Mpc. (*b*) The same comparison if $D = 12.2$ Mpc. The dashed line shows Wevers's (1984) estimate of the sky brightness.

If we divide the surface density Σ_{20} by a suitable mass:light ratio $\Upsilon \approx 5\Upsilon_{\odot}$, and then add a background luminosity density corresponding to $\mu = 21$ mag arcsec^{-2}, we obtain the net photometric profile for NGC 3198 that is shown by the full curve in figure 4*a*.

Wevers (1984) gives F-band photographic photometry of this galaxy. From his data he has subtracted a sky brightness equal to 21.05 mag arcsec^{-2}, but his data are not corrected to face-on, as they must be before they can be compared with the predictions of figure 4*a*. NGC 3198 is inclined at 70°, so I multiply Wevers' brightnesses by cos 70° to allow for the increased area of the galaxy at face-on orientation, and by dex $[0.4 \times 0.215(\sec 70° - 1)] \approx 1.46$ to allow for reduced internal absorption at face-on inclination. The squares in figure 4*a* show the resulting net brightnesses. These clearly fall significantly lower than the values derived from the rotation curve. However, the fact that the photometric points lie under the kinematically derived curve inside two disc scale-lengths (5.36 kpc), suggests either that the distance given by van Albada *et al.*, $D = 9.2$ Mpc, is an underestimate, or that the correction for internal absorption is too small. Figure 4*b* shows that increasing the distance to $D = 12.2$ Mpc lowers the theoretical curve onto the photometric data at small R, leaving a discrepancy between theory and observation that rises to 8% of the sky at 8 kpc, falling to about 4% of sky at the last photometric point ($R = 23$ kpc). In view of this situation one must examine carefully the possibility that Wevers' sky level suffers from systematic errors.

Determining the sky background

I cannot speak with authority in this matter because I have no direct experience of raw photometric data. But reading the accounts of experts in the field has not persuaded me that one would be wise to stake a sum in excess of, say, £100 on the proposition that at 23 kpc Wevers' sky level is less than 4% in error. And it must be emphasized that all discussion of dark matter rests on these and similar photometric data. So any possibility of error must be rigorously examined.

Since the brightness of the night sky is both time and space dependent, the sky level must be determined from the same exposure as the galaxy's brightness. This requirement severely limits the scope of work with currently available CCD detectors, which lack the field of view required to cover a great range of radii in one exposure. Hence the mainstay of this work is still photographic photometry. In outline the most commonly used procedure for estimating the sky level is as follows. (i) An annulus of 'sky' around the galaxy is selected by estimating (usually from the exponential law) from which radius the galaxy's contributions can be neglected. (ii) Stars, plate defects and so forth are eliminated from the data for this annulus by considering the histogram of brightnesses for pixels within the annulus. (iii) A low-order polynomial is fitted to the remaining pixels, and the extrapolation of this polynomial in to the centre of the annulus defines the sky level. Note (i) that the derived brightness of the galaxy is *by construction* zero in the sky annulus, and (ii) that if the galaxy actually contributes non-negligibly to brightnesses in the sky annulus, the fitted polynomial will carry any gradient in the galaxy's contributions into the centre. Hence in this case the derived sky background will be domed towards the centre in just such a way as is required to push the dots in figure 4*b* up towards the theoretical curve. Thus current photometric techniques may be simply incapable of detecting the extended light distributions that are predicted by the hypothesis $\Upsilon \approx$ const.

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

5. Summing up

I undertook to contribute to this meeting very much in the spirit of a barrister who agrees to act for a client who is probably guilty as charged, but is none the less entitled to a fair trial. Yet in working up the case my doubts about the strength of the prosecution's arguments have grown to the degree that my voice would now resonate with conviction as I delivered the following winding-up speech.

Members of the jury, the galaxies stand before you charged with unlawfully possessing unlighted mass. Although the evidence presented by the prosecution that galaxies possess great hoards of mass at large radii is strong only in a very few cases, my clients do not wish to contest the charge of possession. They contest, rather, the charge that the aforesaid matter is unlighted.

We have shown that in their well-observed inner regions, galaxies of all types scrupulously observe the requirement that each unit of mass should be associated with the lawful measure of light. Furthermore, we have shown that the galaxies are law abiding in this respect not from brute necessity, but because observance of this particular law is written into the genetic code that governs the formation of their very tissue. In short, until the emergence of the present charge, the galaxies may be shown to have been citizens of unblemished character who have been ever punctilious in the prompt and exact payment of their light-dues. Will you in these circumstances condemn them while any reasonable doubt exists that they may, after all, be current in all their obligations?

In the preliminary hearings (Knapp & Kormendy 1986) a cornerstone of the prosecution's case was formed by the optical rotation curves of Rubin *et al.* (1982). But today we have heard that a more meticulous examination of the light returns by Kent has so weakened this cornerstone of the prosecution's case that my learned friend Professor van Albada wishes to withdraw it and substitute instead the radio obervations of Bosma (1978) and Begeman (1986), particularly those of NGC 3198. But, members of the jury, you cannot remove one cornerstone and substitute another without undermining our confidence in the stability of the building; will the new cornerstone not soon be crushed by the same weight of photometric evidence as was the old one?

Certainly Begeman's observations indicate that NGC 3198 possesses much more mass beyond about 2.5 disc scale lengths than Wevers' light returns suggest is lawful. But with figure 4*b* we have shown that this conclusion would be nullified by an 8% uncertainty in Wevers' light background. Furthermore, the conventional technique for determining the light background *assumes* that galaxies emit no light at radii comparable to those at which they are currently suspected of possessing mass, and there is a prima facie case that this assumption may seriously compromise the background level derived for smaller radii. Yet the prosecution has presented no evidence to support its implied contention that background level can be determined to the required accuracy, relying instead on the written testimony of experts in photographic photometry and a mass of speculation as to how the galaxies could have concealed matter in weakly interacting particles. Members of the jury, turn a deaf ear to idle gossip about fantastical particles in which it is in principle possible to conceal much of one's substance. Rather let yourselves be guided by your common experience of the worth of expert testimony, and return a verdict of not guilty, or at the very least, of not proven!

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY A MATHEMATICAL, PHYSICAL & ENGINEERING SCIENCES

REFERENCES

Allen, C. W. 1973 *Astrophysical quantities*, p. 248. London: Athlone Press.
Bahcall, J. N. 1984*a* *Astrophys. J.* **276**, 169.
Bahcall, J. N. 1984*b* *Astrophys. J.* **287**, 926.
Begeman, K. 1986 Ph.D. thesis, University of Groningen.
Bosma, A. 1978 The distribution and kinematics of neutral hydrogen in spiral galaxies. Ph.D. thesis, University of Groningen.
Burton, W. B. & Gordon, M. A. 1978 *Astron. Astrophys.* **63**, 7.
de Vaucouleurs, G. & Pence, W. D. 1978 *Astr. J.* **83**, 1163.
Hill, G., Hilditch, R. W. & Barnes, J. V. 1979 *Mon. Not. R. astr. Soc.* **186**, 813.
Knapp, G. R. & Kormendy, J. (eds) 1986 In *Dark matter in the Universe* (*IAU symposium* no. 117) Dordrecht: Reidel.
Knapp, G. R., Tremaine, S. D. & Gunn, J. E. 1978 *Astr. J.* **83**, 1585.
Lucy, L. 1974 *Astr. J.* **79**, 745.
Oort, J. H. 1965 In *Galactic structure* (ed. A. Blaauw & M. Schmidt), pp. 455–509. Chicago University Press.
Rubin, V. C., Ford, W. K., Thonnard, N. & Burstein, D. 1982 *Astrophys. J.* **261**, 439.
Satoh, C. 1980 *Publ. Astron. Soc. Japan* **32**, 41.
van Albada, T. S. Bahcall, J. N., Begeman, K. & Sancisi, R. 1985 *Astrophys. J.* **295**, 305.
Wevers, B. M. H. R. 1984 A study of spiral galaxies. Ph.D. thesis, University of Groningen.

Discussion

R. J. TAYLER (*Astronomy Centre, University of Sussex*). Is there any dynamical reason for believing that an r^{-2} density law might be the correct one for galaxies?

J. J. BINNEY. Numerical simulations of the collapse and violent relaxation of initially sharply bounded and fairly homogeneous protogalaxies yield relaxed systems in which $\rho \sim r^{-4}$ at large r. However, Gunn (1978) has argued that in a critically dense Universe, subsequent infall would add to such a core an envelope in which $\rho \sim r^{-2}$. This argument, however, would seem only to apply to relatively isolated galaxies.

Additional reference

Gunn, J. E. 1978 In *Observational cosmology* (ed. A. Maeder, L. Martinet & G. Tammann), p. 1. Geneva Observatory.