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

Dark Matter near the Sun

J. N. Bahcall

Phil. Trans. R. Soc. Lond. A 1986 **320**, 543-551
doi: 10.1098/rsta.1986.0134

Phil. Trans. R. Soc. Lond. A **320**, 543–551 (1986)
Printed in Great Britain

Dark matter near the Sun

By J. N. Bahcall
Institute for Advanced Study, Princeton, New Jersey 08540, *U.S.A.*

The amount of dark matter in the disc of the Galaxy at the solar position is determined by comparing the observed distributions of tracer stars with the predictions obtained from different assumptions of how the unseen matter is distributed. The major uncertainties, observational and theoretical, are estimated. *For all the observed samples, typical models imply that about half of the mass in the solar vicinity must be in the form of unobserved matter.* The volume density of *unobserved* material near the Sun is about $0.1\ M_{\odot}\ \mathrm{pc}^{-3}$; the corresponding column density is about $30\ M_{\odot}\ \mathrm{pc}^{-2}$ ($1\ \mathrm{pc} \approx 30857 \times 10^{12}\ \mathrm{m}$). This, so far unseen, material must be in a disc with an exponential scale height of less than 0.7 kpc. All the existing observations are consistent with the unseen disc material being in the form of stars not massive enough to burn hydrogen. It is suggested that the unseen material that is required to hold up the rotation curves of galaxies and to satisfy the virial theorem for clusters of galaxies might also be in the form of low-mass stars.

1. Introduction

The main results that I wish to convince you of are listed below:

$$0.5 \leqslant \rho^{(0)}_{\mathrm{unobs}}/\rho^{(0)}_{\mathrm{obs}} \leqslant 1.5 \tag{1}$$

and

$$z_{\mathrm{scale\ height}} \leqslant 0.7\ \mathrm{kpc}. \tag{2}$$

The first equation says that the amount of unobserved material in the vicinity of the Sun is between 0.5 and 1.5 times the already observed material. The second equation says that the exponential scale height of the unobserved material, if it is a single population, must not exceed 0.7 kpc. Thus about half of the matter in the vicinity of the Sun is in the form of unseen disc material which has a scale height of less than 0.7 kpc. I will also argue, in my last remarks, that the available observations are consistent with the dark material in the disc being in the form of stars that are not massive enough to burn hydrogen.

Is the missing material in the outer reaches of galaxies and clusters of galaxies the same as the dark material in the disc? No, if we restrict ourselves to considering the various particle physics solutions that have mainly been discussed at this conference. All of the 'inos' that have been considered are dissipationless and therefore unlikely to gather themselves in a disc, as required by the analysis of the local missing matter.

However, it is possible that all of the dark material consists of low-mass stars that were produced in the first stages of galaxy formation. The 'conspiracy' between dark and seen material (see, for example, Bahcall & Casertano 1985) which makes rotation curves relatively featureless has been discussed a number of times in this symposium. This conspiracy is not particularly sinister or surprising if the unseen material is made of the same stuff as the seen material. Low-mass stars could be in an approximately spherical distribution around galaxies and, if their initial mass function were appropriately weighted towards small stars, they would not produce too many heavy elements.

Before we get down to the justification of the main results, I want to remind you of some of the history of this subject because it may be of significance in this connection. Oort's (1932, 1960) early studies of the total amount of matter in the solar vicinity led to what may have been the first astronomical suggestion of a large 'missing mass'. Is it not just possible that the solution of this first missing matter problem contains the key to understanding, as well, dark matter in galaxy halos and clusters of galaxies?

2. The method

The method of weighing the matter in the local neighbourhood that I have used, and which Oort pioneered, can be summarized as follows. A detailed model of the observed matter (in stars, gas and clouds) is constructed from all the available observations. In addition, the density distribution and velocity dispersion of a set of tracer stars perpendicular to the galactic plane is taken from published measurements. Theoretical models are then computed for the expected distribution of tracer stars in different gravitational potentials (mass distributions). The amount of matter that is actually present in the Galaxy is determined by comparing the observed and computed distributions.

The problem is similar to computing the distribution of an isothermal atmosphere (because for the tracer stars of interest the velocity dispersion changes much more slowly with height above the plane than does the density). Clearly, the more matter there is close to the plane, the more quickly will the density fall off with height above the plane.

The availability of modern computers has made possible important improvements in the theoretical analysis of this problem at the same time that better observational samples of tracer stars have been obtained. I have taken advantage of these developments to sharpen the determinations of the total amount of matter in the solar vicinity, using more realistic galaxy models and more accurate theoretical solutions. I have solved numerically the combined Poisson and Vlasov equations for the gravitational potential of galaxy models consisting of realistically large numbers of individual isothermal disc components in the presence of a massive unseen halo.

Most previous calculations were carried out without requiring self consistency between the Poisson and Vlasov equations. For example, in Oort's work the equations were solved separately. In his first discussion he did not solve Poisson's equation and therefore obtained, at large distances above the plane, negative mass densities. In his 1960 investigation Oort took the mass density on the right hand side of Poisson's equation to be fixed, not responding to the gravitational potential.

In the solutions that I will discuss, the distribution functions that solve Vlasov's equation for the observed matter and the tracer stars also depend on the potential that appears in Poisson's equation and generate, through their associated densities, the mass densities in Poisson's equation. I have carried out the calculations with different assumptions about the unseen matter and have compared the results with the observed number densities of F dwarfs and K giants against height above the plane, assuming that the F dwarfs and K giants are reasonably faithful tracers of the total gravitational potential. Because the solutions are obtained with the aid of a computer, I can make more quantitative estimates of the errors by varying all of the parameters and by trying many different models.

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

Incidentally, the work of Oort and other previous investigators referred only to the equivalent of (1) above. The derivation of (2) requires combining the studies of the motion perpendicular to the plane with knowledge of the galaxy rotation curve.

3. The input data

The relative amounts of the observed mass components and their velocity dispersions (i.e. temperatures) that were derived, from data from many sources, by Bahcall & Soneira (1980) are summarized in table 1. A similar model was derived earlier by Hill *et al.* (1979). These two models are often referred to, respectively, as the B.S. and the H.H.B. galaxy models. The B.S. model contains many observed disc components (typically 14) whose characteristics are determined by local measurements: a population II spheroid inferred from faint-star counts; different models for the unobserved disc components and an unseen massive halo whose normalization is fixed by the solar rotation velocity. The mass fractions are defined in terms of the total *observed* mass density (in stars, gas and dust), i.e.

$$A_i = \rho_i(0)/\rho_{obs}(0). \tag{3}$$

The total *observed* mass in the B.S. model is 0.096 $M_\odot \mathrm{pc}^{-3}$†.

Table 1. Galaxy model: B.S.

component	local density $M_\odot \mathrm{pc}^{-3}$
main-sequence stars (11 components)	0.044
subgiants and giants	0.0015
white dwarfs	0.005
interstellar matter	0.045
spheroid	0.0001
unseen halo	0.01

I use the difference between the results obtained with the B.S. and the H.H.B. galaxy models as one measure of the uncertainty. The two models are similar because the luminosity function of the disc stars is reasonably well determined (see Wielen 1974) over much of its range. The B.S. and H.H.B. models mainly differ in the mass density assigned to white dwarfs and to the interstellar matter. In both cases I have made use of more recent determinations. For example, fewer white dwarfs are observed at faint absolute magnitudes than had been expected on the basis of earlier theoretical estimates. I have used in the B.S. model the observed number density (Green 1980; Liebert *et al.* 1979) down to $M_V = 17.2$ and a white dwarf mass of 0.6 $M_\odot$. I have also adopted the value for interstellar matter density that has been estimated by Spitzer (1978), which is consistent with the recent value inferred by Sanders *et al.* (1984). This value is rather larger than the interstellar matter density that was used by H.H.B.

In my discussion, 'dark' matter will not be synonymous with 'exotic' or with 'unobservable'. I shall use the term dark matter to refer to any material that has not *yet* been observed.

Previous theoretical studies of the total amount of matter in the vicinity of the Sun have been limited to simplified galaxy models with one or, at most, a few disc components and no spherical component. The previous solutions were also limited either by what was tractable analytically or by assuming a numerical form for the total matter density that was independent of the

† 1 pc ≈ 30857 × 10^{12} m.

potential. As I have access to a VAX computer, I have calculated numerical models with many different sets of input data and several assumptions about the unseen material is distributed. I estimate the major uncertainties in the determination of the distribution of unseen matter by comparing an extensive collection of theoretical models with the available data.

4. The simplest model for the unseen material

Because we have not yet observed the unseen material, we do not know how it is distributed. Therefore we have to try different models for the unseen material to see how the results depend upon our assumptions.

There is one model which is uniquely simple and is characterized by only one parameter, the overall scale factor, P, between observed and unobserved material. In this illustrative model, the unobserved mass density in every component, i, is proportional to the observed mass density in the same component,

$$B_i \equiv P \times A_i, \tag{4}$$

and the unobserved and observed velocity dispersions for the ith component are equal. Of course, this is only one of the many different models that have been explored.

Figure 1 is a chi-by-eye illustration of why one needs missing matter in the disc. I compare in this figure the measured star densities of Hill *et al.* (1979) with a sequence of models computed assuming that the scale factor, $P = 0.0$, 0.3, 0.5, 0.75, 0.97, 1.2, 1.5 and 2.0. You can judge for yourself the improvement in the agreement between model and observation as the amount of material is increased from no unobserved material ($P = 0.0$), through the best fit ($P = 0.97$), to a worsening of the fit at large ratios of unobserved to observed matter (up to an unacceptable $P = 2.0$). For small values of P, the observed distribution of F stars falls off more rapidly than does the calculated distribution. Therefore, we have to add additional unseen matter to pull down the calculated curve. A formal treatment (Bahcall 19842*a*) gives for this case $P = 0.97 \pm 0.23$. Does that agree with our chi-by-eye assessment of the uncertainty? Incidentally, the flatness of the observed distribution within the first 40 pc (the first three data points in figure 1) is an artifact of the way that Hill *et al.* reduced their data and does not reflect any real observational constraint on the shape of the distribution at small heights above the plane.

The agreement between the number densities for the Oort (1960) and the Upgren (1962) samples of K giants is shown in figure 2. The best fit for the Oort data, again with the simple proportional model, is shown in figure 3.

5. Other models and equation (1)

I have explored many possible models for the distribution of unobserved material. I have calculated, for example, models in which the unobserved material has a small velocity dispersion (like the interstellar material), has a distribution like the older stars (e.g. like the white dwarfs or K giants), is distributed like all the observed stars (ignoring the interstellar material), or has the maximum scaleheight consistent with the galaxy rotation curve.

Table 2 gives the ratio of unobserved to observed mass density for 28 detailed models (see Bahcall 1984*b* for a description of these models) that fit the observed distribution of K giants.

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

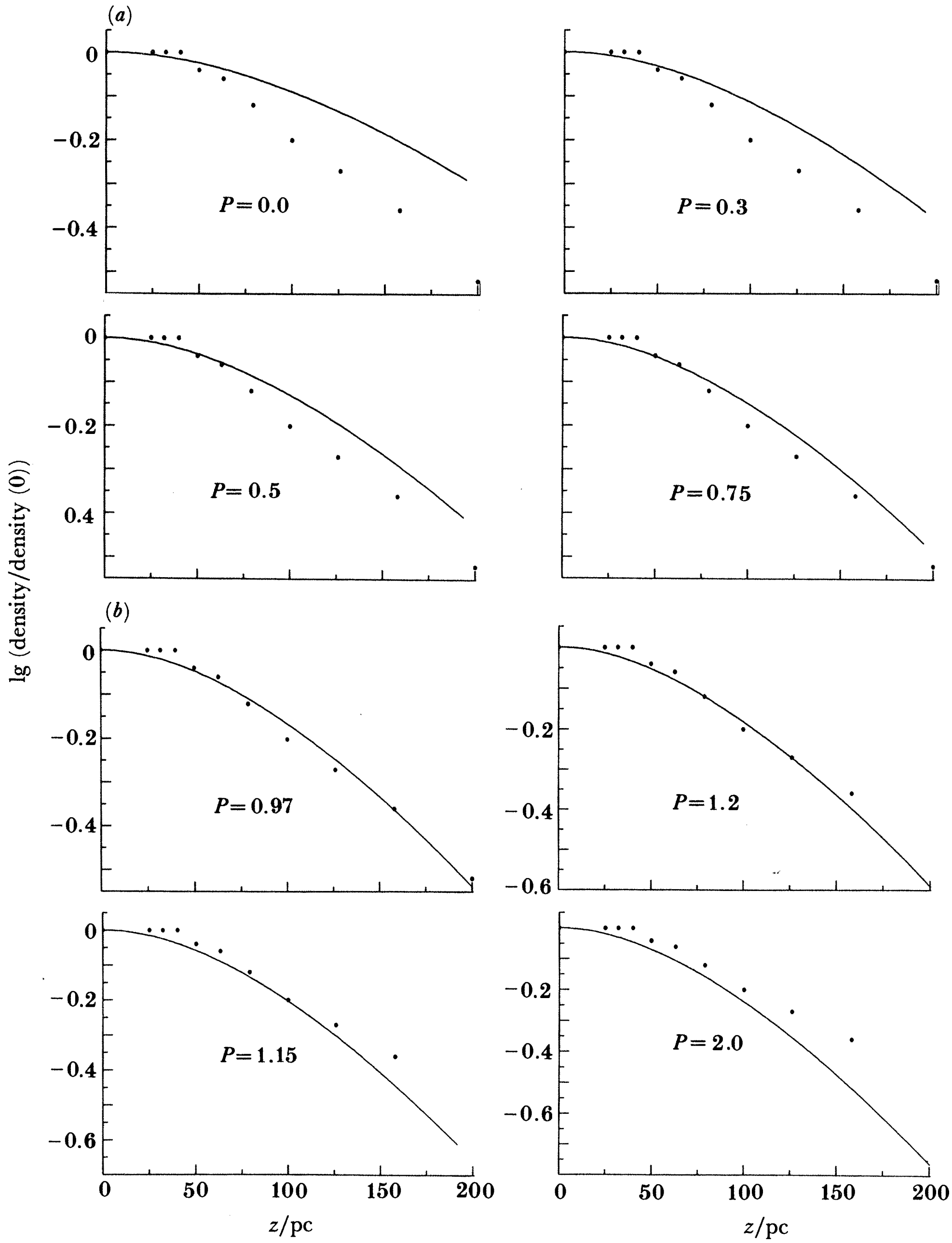

FIGURE 1. Comparison of measured against computed number densities of F stars. The measured densities are taken from the work of Hill *et al.* (1979). The sequence of theoretical models is described in the text and in rows 1–8 of table 2. The mass in unobserved material is assumed to be proportional to the mass in observed material, stellar and interstellar, with proportionalty constant *P*.

The models represent numerical solutions of the combined Poisson–Vlasov equation for different input parameters, as well as for several assumptions about the distribution of the unobserved disc material. There are separate columns referring to the observed K-giant samples of Oort (1960) and to the Upgren (1962) K-giant density distributions. For both the volume and the column density, the typical best-fit model has, for the Oort densities, about equal amounts of unobserved and observed material. For the Upgren densities, the typical

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

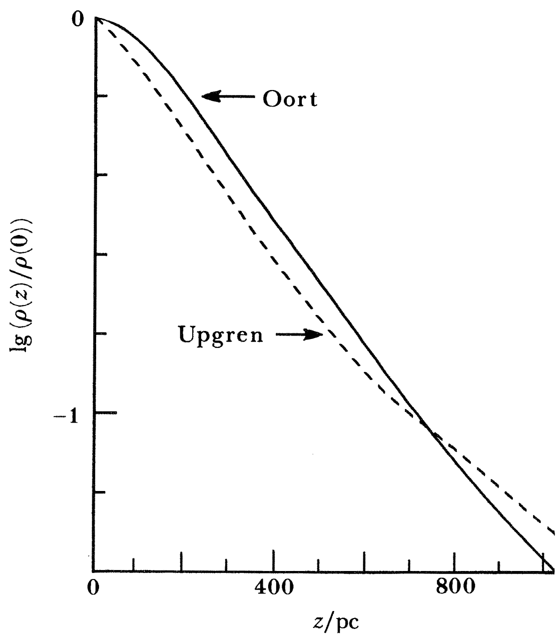

FIGURE 2. The comparison of the Oort (1960) and Upgren (1962) density distribution with the average visual magnitude and absorption adopted in Bahcall (1984*b*).

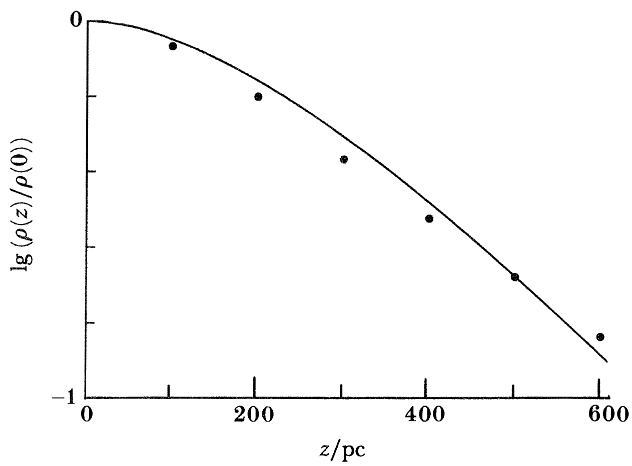

FIGURE 3. The best-fit model for the Oort (1960) K-giant data with the simple scale model defined in §3 of this paper ($P = 1.1$).

best-fit model has about 40 % more unobserved than observed matter. *These averages are only illustrative because at most one of the models considered for the distribution of unseen matter can be correct.* Similar results are obtained by comparing theoretical models to the observed sample of F dwarfs (Bahcall 1984*a*).

I conclude that *a typical best-fit model implies that about half of the disc material at the solar position has not yet been observed.* This conclusion, which is summarized in (1), is in qualitative agreement with the previous major studies (see, for example, Oort 1932, 1960; Hill 1960; Woolley & Stewart 1967; Lacarrieu 1971; Hill *et al.* 1979), although I find a larger ratio of unobserved to observed matter than in some of the earlier analyses. The present investigation establishes more firmly and specifically the existence of unobserved disc material. The added confidence

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

Table 2. Ratio of unobserved to observed disc material

	Oort densities		Upgren densities	
row[1] (1)	$\frac{\rho_{unobs}(0)}{\rho_{obs}(0)}$ (2)	$\frac{\sigma_{unobs}}{\sigma_{obs}}$ (3)	$\frac{\rho_{unobs}(0)}{\rho_{obs}(0)}$ (4)	$\frac{\sigma_{unobs}}{\sigma_{obs}}$ (5)
1	1.1	1.1	1.5	1.6
2	1.6	1.6	2.1	2.1
3	0.6	0.6	1.0	1.0
4	0.9	0.9	1.3	1.4
5	1.3	1.3	1.8	1.8
6	1.3	1.1	1.8	1.6
7	0.6	1.2	0.8	1.6
8	0.7	0.7	1.0	1.0
9	0.4	0.7	0.5	1.0
10	2.4	0.5	2.6	0.5
11	1.5	0.3	2.2	0.5
12	0.6	2.5	0.7	3.2
13	1.1	1.1	1.5	1.6
14	1.5	1.5	2.0	2.0
average	1.1	1.1	1.5	1.5

[1] Disc luminosity functions and velocity dispersions from Wielen (1974).

in the results arises because: (1) more realistic galaxy models are used; (2) the Poisson and Vlasov equations are solved self consistently; (3) improved (and more homogeneous) observational data are utilized and (4) many theoretical models are compared with the observations in order to estimate the uncertainties.

6. But...

I do not want to sound too satisfied, however. There is no modern data sample of K giants; the samples that I have been forced to use are a quarter of a century old! The stars are very bright (apparent magnitudes less than 10) so that it would be very easy to get a much improved sample with modern techniques, by using spectroscopic observations to assure that the population was homogeneous with height above the plane. The velocity dispersions of both the K giants and the F dwarfs could be improved with modern radial velocity techniques. Finally, the absolute magnitude of the tracer stars should be redetermined using Hipparcos as well as the Yale parallax catalogue, soon to be published.

The largest identifiable source of uncertainty in the Oort limit is the unknown form of the distribution of unseen matter (see the last row of table 9 of Bahcall 1984*b*). In the future, it should be possible to constrain sharply the distribution of unseen matter by requiring consistency with observations of several carefully selected samples of tracer stars with different scaleheights.

7. The rotation curve and equation (2)

The unseen material must be mostly in a disc form, i.e. be dissipational. If all of the material were in a relatively round halo, then the rotation velocity at the solar position would have to be as large as 500 km s^{-1}. For a given local volume density of unseen mas, the total amount of mass required in a round halo is larger than the amount of mass needed in a disc by about the ratio of the galactocentric distance of the Sun to the disc scale height, i.e. by more than an order of magnitude. The largest exponential scale height of the unseen disc material that is consistent with the solar rotation velocity is 0.7 kpc (see row 12 of tables 5 and 6 of Bahcall 1984*b*). I determined this value by making a succession of models in which the unseen material had a progressively larger vertical velocity dispersion. For each model I required that the predicted distributions of tracer stars fit the observations of F dwarfs and K giants and also be consistent with the observed (220 km s^{-1}) rotation velocity at the solar galactocentric position. The maximum allowed vertical velocity dispersion is 40 km s^{-1}.

8. What is it?

I think that the most plausible form in which the unseen disc matter could reside is that of faint stars. This is a conservative solution because we do observe stars with a total local mass density that is within a factor of three of what is required to explain the disc missing matter. Moreover, it is perfectly plausible that the stellar mass function is everywhere continuous but that most of the stars around today happen to lie below the minimum mass for hydrogen burning. We would be lucky if it were otherwise but we have no justification for imposing our good fortune as a necessary condition for an acceptable description of the observations. All of the available observations are consistent with the conservative interpretation that the missing disc mass is mostly in brown dwarfs. The turnover of the observed luminosity function in the range of absolute visual magnitudes between 12 and 14 may simply reflect the steepening of the bolometric correction in this region (D'Antona & Mazzitelli 1983).

What does one have to do to order to go from a theoretical mass density, say 0.1 $M_\odot$ per cubic parsec, to a predicted number of stars per square degree on the sky? We need to know, first, a mass function as it depends upon galactic age, $N(m, t)$. Theory really gives us no reliable handle on this problem at present. Most of the stars could be born with typical masses of 0.02 $M_\odot$, in which case we are only observing the far tail of the distribution. *In order to proceed, we need to have faith that the missing mass will be in the observable region.* If we assume that we know the shape of the mass function, and it is favourable, we need next to relate mass to total luminosity and galactic age. Fortunately, this step can be done relatively reliably with existing theoretical models. The next stage, however, is much more difficult. What are the broad band colours of an object with a specified total luminosity and mass? Theorists are not willing to propose answers to this question because it depends upon unknown opacities, the nature and amount of huge grains, and the undetermined surface composition. Observers quite correctly say they cannot be asked to establish the luminosity–colour characteristics of a population that has not yet been identified.

One can express this difficulty of confronting expectation and observation in a different way. Observers usually report their results in terms of a luminosity function, for example, $\mathrm{d}N/\mathrm{d}M_V$. This is not a directly measured quantity, but suppose, for simplicity, it were. What do we have to know to convert an 'observed' luminosity function into the theoretically desired mass function? The relation between the luminosity function and the mass function is given by the chain rule:

$$\mathrm{d}N/\mathrm{d}M_V = (\mathrm{d}N/\mathrm{d}m) \times (\mathrm{d}m/\mathrm{d}L) \times \mathrm{d}L/\mathrm{d}M_V. \tag{5}$$

Here N is the measured number of stars, m is the mass, L the total luminosity, V is the observing band and M_V is the absolute magnitude. For any colour, $(\mathrm{d}m/\mathrm{d}L)$ must be very close to zero near the transition region between nuclear burning (M dwarfs) and gravitationally supported objects (brown dwarfs). The luminosity changes by more than two orders of magnitude whereas the mass changes by less than 10%. Thus we have to divide by a very tiny quantity, $(\mathrm{d}m/\mathrm{d}L)$, in order to convert the measured values into a mass function. What is more $\mathrm{d}m/\mathrm{d}L$ is not very well determined just around the transition region where, among other things, the age of brown dwarfs strongly affects their luminosity. For simplicity, I have suppressed the age dependence in (5). But, the biggest uncertainty in determining $(\mathrm{d}N/\mathrm{d}m)$ comes from the last term in (5), the colour relation. Theorists and observers alike agree they do not know how to determine reliably $(\mathrm{d}L/\mathrm{d}M_V)$.

I conclude that it will not be easy to infer a mass function of brown dwarfs from observations obtained with current techniques. I do not believe existing observations strongly constrain the mass function below 0.1 $M_\odot$.

Note that the constraint that the exponential scale height of the unseen disc material must be less than 0.7 kpc is easily satisfied if the local unseen material is in the form of faint stars, but is not easy to understand if the material is in the form of (dissipationless) elementary particles.

Suppose that the unseen disc material is in the form of low-mass stars. What does this imply for the unseen material that holds up the rotation curves of galaxies and is required in order to satisfy the virial theorem in clusters of galaxies? Logically, nothing. There could be at least two kinds of dark matter. However, I think that it is possible that nearly all of the unseen material is in the form of low mass stars, whether in the galaxy disc, the outer reaches of galaxies or in clusters of galaxies. At the very least, this is a well-defined and conservative hypothesis that ought to be tested in every way possible.

This work was supported in part by the National Science Foundation grants PHY-8217352 and by NAS8-32902.

References

Bahcall, J. N. 1984*a* *Astrophys. J.* **276**, 169.
Bahcall, J. N. 1984*b* *Astrophys. J.* **287**, 926.
Bahcall, J. N. & Casertano, S. 1985 *Astrophys. J. Lett.* **293**, L7.
Bahcall, J. N. & Soneira, R. M. 1980 *Astrophys. J. Suppl.* **44**, 73.
D'Antona, F. & Mazzitelli, I. 1983 *Astron. Astrophys.* **127**, 149.
Green, R. F. 1980 *Astrophys. J.* **238**, 685.
Hill, E. R. 1960 *Bull. astr. Insts, Neth.* **15**, 1.
Hill, G., Hilditch, R. W. & Barnes, J. V. 1979 *Mon. Not. R. astr. Soc.* **186**, 813.
Lacarrieu, C. T. 1971 *Astron. Astrophys.* **14**, 95.
Liebert, J., Dahn, C. C., Gresham, M. & Strittmatter, P. A. 1979 *Astrophys. J.* **233**, 226.
Milgrom, M. A. 1983 *Astrophys. J.* **270**, 371.
Oort, J. H. 1932 *Bull. astr. Insts, Neth.* **6**, 249.
Oort, J. H. 1960 *Bull. astr. Insts, Neth.* **15**, 45.
Sanders, D. B., Solomon, P. M. & Scoville, N. Z. 1984 *Astrophys. J.* **276**, 182.
Spitzer, L. 1978 *Physical processes in the interstellar medium.* New York: Wiley.
Upgren, A. R. 1962 *Astr. J.* **67**, 37.
Wielen, R. 1974 In *Highlights of astronomy* (ed. G. Contopoulos), vol. 3, p. 395. Dordrecht: D. Reidel.
Woolley, R. & Stewart, J. M. 1967 *Mon. Not. R. astr. Soc.* **136**, 329.

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