BIASED GALAXY FORMATION AND DARK MATTER

Martin J. Rees
Institute of Astronomy
Madingley Road, Cambridge CB3 OHA
England

ABSTRACT. The clustering properties of galaxies are incompatible with a cosmological model with $\Omega = 1$ unless the formation process for bright galaxies is 'biased' in the sense that their resultant distribution exaggerates the amplitude of large-scale inhomogeneities in the overall mass distribution. The mechanisms for such biasing are intimately connected with the nature of the dark matter. Various possibilities are summarised here.

1. INTRODUCTION: SOME COMMENTS ON DARK MATTER

The relationship between the distributions of luminosity and matter in the Universe is bound to remain mysterious until we know what the gravitationally-dominant constituent of the Universe actually is. The quantitative details of galaxy formation, and any disparity that may emerge between the way light and gravitating matter are distributed, plainly depend on the nature of the dark matter. If this matter were non-baryonic, we would have our choice amongst the fauna of the 'Turner zoo' (see his accompanying paper). Three lines of investigation should then help to discriminate among the candidate particles:

(i) Firmer knowledge of what particles should exist in the early Universe, and of their masses and cross-sections, may allow physicists to predict, on the basis of standard Friedman cosmology, that one particular species left over from the big bang should survive in sufficient numbers to be dynamically dominant.

(ii) The cosmogonic consequences of various different possibilities can be explored, especially via computer simulations of the development of bound structures from small-amplitude initial perturbations. This approach cannot pin down exactly what the particles are, but may decide between candidates which could be treated as 'cold', in the sense that their thermal motions would be too slow for free-streaming effects to erase fluctuations on any astronomically-interesting scales, and those that were 'hotter'. Such studies may also yield evidence for or against particles which decay on timescales of 10^9 - 10^{10} yrs.

J. Audouze et al. (eds.), Large Scale Structures of the Universe, 437–446.

https://doi.org/10.1017/S0074180900136344 Published online by Cambridge University Press

(iii) Direct experimental searches might reveal the particles pervading our own galactic halo. Should such searches succeed, and a background of particles with velocities ~ 300 km s^{-1} be found, the payoff for physics and astronomy could be at least as important as the discovery of the microwave background. These are therefore perhaps the most worthwhile of high-risk laboratory experiments.

Biasing can arise in two ways:
(a) from segregation of baryonic and non-baryonic matter, and/or
(b) as an outcome of spatial variations in the M/L for baryonic matter, resulting from differing efficiencies for conversion of gas into stars, or from differences in the IMF of the stars that do form. (Of course only the second of these options would survive if the Universe were baryon-dominated.)

Baryonic dark matter is undeniably a duller option from any physicist's viewpoint. It therefore tends to get rather short shrift, and I should like to redress the balance by venturing one remark urging the inherent plausibility of the dark matter being in stars (or stellar remnants) with very high M/L. At first sight, this idea may seem difficult to reconcile with the claimed sharp demarcation between ordinary matter – stellar populations (whose M/L lies in a well defined range depending on the IMF, the history of star formation, etc.) and gas – and 'dark' matter with M/L > 1000. The big difference between the M/L for these two components, in contrast to the relatively small spread in M/L for the stellar populations actually observed in galaxies, is sometimes adduced as support for non-baryonic dark matter. If the matter is nonetheless postulated to be baryonic, a very exotic mode of pregalactic star formation is customarily invoked, on the grounds that the 'dark' stars must have formed in an environment far removed from the conditions prevailing anywhere within the observed galaxy (where 'normal' star formation always seems to yield M/L somewhere in the standard range).

I think this difficulty may be exaggerated: indeed once we accept that there may, even within our own Galaxy, be more than one mode of star formation, and that stars forming in different sites may have very different mean masses, a relatively sharp demarcation between ordinary populations with $M/L \simeq 1$ and dark populations for $M/L \gg 1$ seems quite natural.

A star-forming region can be very bright, with a very low value of M/L, when a starburst occurs, but any such phase would be brief. The mean luminosity in any star-forming region is constrained by the nuclear energy available. The lives of blue stars are shorter than cosmological timescales - shorter even than the likely timescale for the most rapid phases of galactic evolution. Therefore a more relevant quantity is the M/L that could be maintained over 10^{10} years by steady star formation – whether this be 10^3 generations of O stars, a few hundred generations of B stars, a single generation of solar-type stars, or very low mass stars. If $\langle L \rangle$ is defined as the luminosity averaged over a Hubble time, then $\langle M/L \rangle$ cannot be significantly below unity (in solar units); moreover this average value is insensitive to the IMF provided only that most of the mass goes into stars somewhere in the range 1-100$M_{\odot}$ (i.e. into stable stars that use up most of their fuel within a Hubble time). However, for an IMF dominated by very low mass stars M/L rises very steeply: the stellar luminosities depend on mass roughly like $M^{3.5}$ and even more steeply below $0.1M_{\odot}$, the threshold for H-burning.

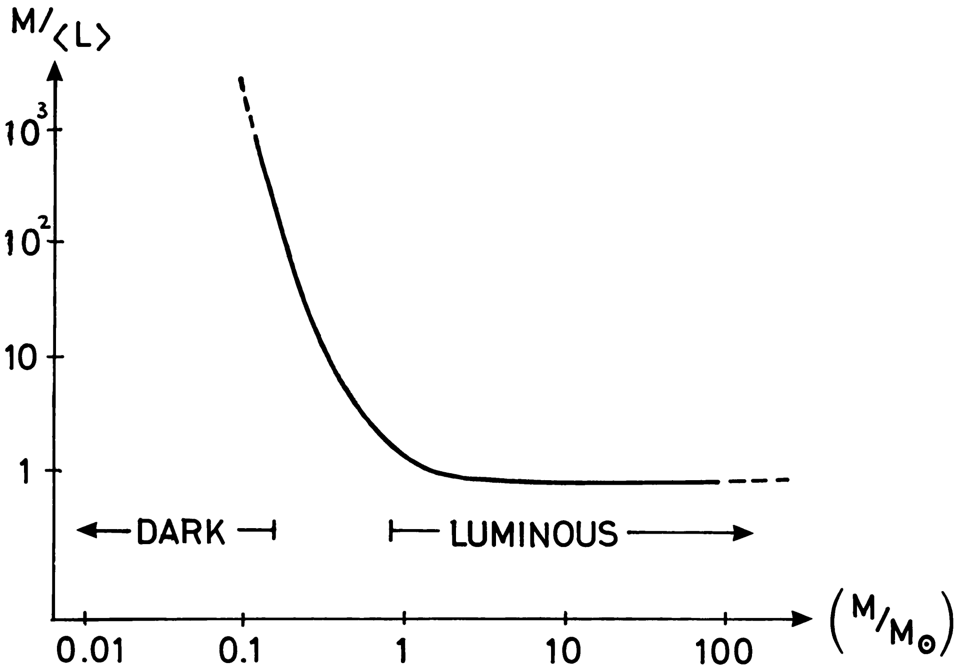

Figure 1. If the stellar IMF was to vary widely, this diagram shows schematically that the mean M/L averaged over a Hubble time is bimodal, being either $\gtrsim 10^3$ or ~ 1, except for ⟨M⟩ in a transitional range 0.2 - 0.8 $M_{\odot}$ (which, from a perspective where the IMF varies widely, could be relatively sharp).

This is depicted schematically in Figure 1. The results, which are trivial and standard, could be calculated for any IMF, which may be different at each stage in galactic evolution and for different modes of star formation. But insofar as we can crudely characterise the IMF by ⟨M⟩, the simple message of Figure 1 is that a wide range of IMFs yield M/L in the 'standard' range, but there is another wide range where the stellar population would have an M/L so high that it would be categorised as dark matter. The intermediate transition region, with M/L between (say) 5 and a few hundred is, in this perspective, a relatively narrow and improbable one. This is merely a rather trivial consequence of adopting the view that there is nothing magic or universal about the local IMF, and that $\langle M \rangle$ may indeed range widely.

In the widely-studied models of galactic evolution where star formation was in an initial burst and/or proceeded at a steadily diminishing rate (e.g. Bruzual 1983), the present luminosity is lower than its time-averaged value; in particular, the remnants of high-mass stars all of which formed early in galactic history could now constitute a dark population. But if there are several modes of star formation, the low-mass mode could contribute dark matter even if it were prolonged: it need not be relegated to an early epoch. (There are, indeed, independent indications that this may even now be the dominant mode within cooling flows, for instance).

2. WHY LUMINOUS GALAXIES NEED NOT TRACE MASS: A CLASSIFICATION OF BIASING SCHEMES

The idea that there may be a large-scale segregation between 'light' and 'mass' has gained wide coverage in the last few years (for a discussion, in an earlier IAU symposium, of how this might arise see Rees (1983)). Avishai Dekel and I recently reviewed some physical ideas on biasing in a Nature article (1987); we emphasised that biasing was essential if $\Omega = 1$, and that there are so many possibilities that it would be astonishing if it did not occur at all. The challenge is to decide among a range of options. In the present written text I shall therefore not repeat the details of points addressed in that paper, but just classify and enumerate various possibilities, mentioning some observational tests that may be relevant.

2.1 Large-scale segregation of baryonic and non-baryonic components

We could clearly not infer the overall distribution of matter from galaxies if the universe were dominated by a non-baryonic component, and the baryons did not themselves trace mass, even on scales $\sim$10 Mpc. Some authors have speculated about explosions sufficiently violent to push matter over distances $\gtrsim$ 10 Mpc (see Ostriker's review in these proceedings). The dark matter would not be completely unaffected, as it would feel the gravity of the displaced baryons, but because of its greater mass it would not all be carried along and would remain relatively smooth. Consequences of these explosions could be a low gas density in voids, bubbles in the galactic distribution, etc.; the gas would be at a high temperature unless the shocks happened early enough that Compton cooling on the microwave background was effective. Another (currently less popular) possibility in this category is that there are primordial n_b/n_γ fluctuations. Were this the case, the

https://doi.org/10.1017/S0074180900136344 Published online by Cambridge University Press

galaxies would obviously form preferentially where the baryons were concentrated. There could be large voids without any of the associated deviations from Hubble flow that would be a concomitant of either gravity or blast waves.

If the baryons are not segregated from a cosmically-dominant non-baryonic component, then we are left with processes that modify the M/L of the baryons — either the fraction that turns into stars, or the IMF of those stars.

2.2 Non-gaussian galactic-scale perturbations

Strings (even the non-conducting variety) would gravitationally-enhance the density in their vicinity, leaving the rest of the baryon distribution relatively smooth. The spatial distribution of the resultant galaxies would then reflect that of the string loops that 'seeded' their formation. This hypothesis, as discussed in Turok's paper, makes distinctive predictions about the correlation function. Strings with the requisite mass per unit length may be detectable via gravitational lensing.

In the adiabatic 'pancake' model, primordial perturbations below the scales of clusters would be damped out. Perturbations on galactic scale would then be secondary, and would not necessarily be Gaussian. (The same statement holds for the inhomogeneities within the shells resulting from large-scale explosions).

If for any reason initial fluctuations had non-random phases — with galaxy formation being triggered only in rough rather than calm patches — the resultant galaxy distribution would depend on the details of the perturbations, rather than tracking the baryon density.

2.3 Biasing related to high peaks of a random-phase gaussian fluctuation field

The peaks rising above some high threshold display enhanced clustering provided that the fluctuation field falls off steeply with amplitude: high peaks occur with enhanced probability in the crests rather than the troughs of a large-scale fluctuation mode. This situation lends itself to a quantitative treatment for gaussian fluctuations, and has received a good deal of attention (Kaiser 1984, Bardeen et al. 1986).

There are various contexts in which this can lead to biasing.

Autonomous or 'natural' biasing. High-amplitude perturbations may convert more efficiently into luminous galaxies:

(a) The high amplitude peaks of a given M yield halos whose higher ρ and higher velocity dispersion V_c may promote more efficient retention of gas (Dekel and Silk 1986). Moreover, because they turn around at earlier times, there is the possibility of enhanced Compton cooling on the microwave background.

(b) Detailed N-body simulations carried out for the specific case of the cold dark matter (CDM) model are reported in Frenk's contribution to this symposium (see also Frenk et al. 1986, White et al. 1987a, b). Of special relevance to biasing is that the high-V_c halos are themselves more clustered than the dark matter in general: this is essentially because the formation of such systems is more likely in overdense regions that mimic a background cosmological model with $\Omega > 1$. It is not obvious how individual luminous galaxies should be identified in these simulations. It is unclear to what extent the largest 'halos' (i.e. recently virialised aggregates of dark matter) represent single big galaxies, rather than groups of galaxies that formed at (say) $z = 2$ and whose luminous cores preserved separate identities even though their halos merged. A specific prescription has recently been proposed by White et al. (1987b). Nevertheless, this important work already suggests that the non-dissipative clustering process may in itself account for the biasing which, if $\Omega = 1$, is obligatory in the CDM model. Although any complete picture of galaxy formation must incorporate gas dynamics, feedback etc, it would be a fortunate simplification if these 'messy' processes were of secondary importance for the biasing.

A general feature of these models is that the biasing should be less evident in the distribution of the smaller halos. If the clouds that cause Lyman α absorption lines in quasar spectra were caused by gas gravitationally bound in minihalos (Rees 1986a), these may display less clustering than the mass in general, because such small and loosely bound systems would be more likely to escape mergers (and consequent incorporation in systems large enough to form galaxies) in underdense regions.

Negative feedback from the first galaxies? For the distribution of galaxies to display the enhanced clustering expected for high-amplitude peaks, something must have prevented lower-amplitude peaks from themselves turning into similar galaxies, thereby neutralising the effect. One possibility is negative feedback from the first galaxies. Since this could occur irrespective of whether or not the fluctuations are gaussian, it is logical to treat it, along with other feedback processes, in a separate section.

3. FEEDBACK PROCESSES

The feedback mechanisms that might cause galaxies to display enhanced clustering compared to the baryons in general could be of two kinds:

3.1 Local positive feedback

The existence of a galaxy may promote an 'epidemic' of galaxy formation in its neighbourhood. This is a feature of the explosive model, but could be important even if the energy output from the first galaxies were too low to generate the large-scale explosions which are the most distinctive feature of that model. If halos were non-baryonic, one might then expect the 'secondary' galaxies to lack such massive dark halos — at least, there is no reason why the ratio of baryonic and non-baryonic matter should be the same in all galaxies.

3.2 Non-local negative feedback

A variety of processes whereby the first galaxies might 'choke off' further galaxy formation were presented by Rees (1985):
(a) Baryons may be inhibited from accumulating in the more typical (later-forming) galactic mass potential wells. This could be because the gas is heated by the first galaxies, or by their active nuclei, to 10^6 K, so that the Jeans mass exceeds a galactic mass (note that photoionization cannot achieve this effect, since it never raises the gas temperature much above 10^4 K). Alternatively, the first galaxies may generate a universal pressure of cosmic rays or trapped radiation which inhibits collapse of later protogalaxies, or through its gradient pushes the baryons relative to the dark matter so that they do not concentrate in dark matter halos. Such effects would tend to be more effective for smaller galaxies with shallow potential wells.
(b) Even if nothing can prevent baryons from accumulating in potential wells, the efficiency of star formation, or the shape of the IMF, may be modified so that the resulting M/L is reduced. The factors controlling the IMF are too poorly understood for this to be more than a speculative suggestion. One possibility is that the background UV flux changes the abundance of H_2, which is an important coolant in metal-poor systems (e.g. Silk 1985). Were any such effect important, one would expect 'failed' galaxies — halos with a high M/L — to exist.

3.3 Local negative feedback

In some cosmogonic schemes, the luminous galaxies have to be <u>less</u> clumped than the matter in general. For instance, in the neutrino-dominated model the amplitude of the dark matter fluctuations on 10 - 20 Mpc scales must be larger than is revealed by the galaxy distribution in order that the first pancakes collapse early enough to account for the existence of high-redshift quasars. To take another example, Peebles (1986, and these proceedings) emphasises that galaxy formation induced by strings whose number-density falls off only as a power law rather than exponentially should lead to a few 'megagalaxies' with larger V_c than is ever observed in any actual individual galaxy. In both these cases, a possible conjecture would be that the less efficient cooling of a very hot gas inhibits star formation, or else (if star formation does occur) the stars form in the low-mass mode inferred to prevail in cooling flows. We cannot exclude the existence of a few megagalaxies if they have as much amorphous dark matter as a cluster of galaxies, but (unlike a cluster) contain no individual galaxies.

3.4 Positive or negative feedback?

It may be worth spelling out a general requirement for negative feedback to be non-local rather than local. Obviously the influence of the first bound systems must propagate fast. The time-lag between the turnaround of bound protogalaxies in an incipient cluster (where

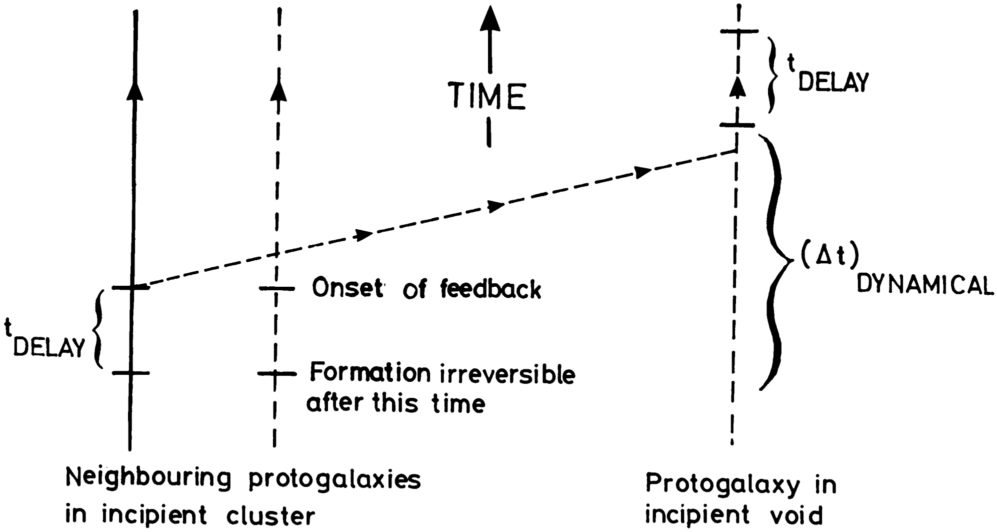

Figure 2. Negative feedback from the galaxies in a proto-cluster could in principle inhibit the later formation of galaxies in an incipient void, thereby creating large-scale biasing (or 'contrast enhancement') in the galaxy distribution. Contrariwise, the negative feedback from an early-forming galaxy could quench the formation of close neighbours, yielding the opposite result. The first of these options requires rapid propagation of the feedback, and also a time-lag t_{delay} between the stage when a galaxy has irrevocably 'formed' (in the sense that it cannot thereafter be quenched by any external influence) and the stage when it starts to exert negative feedback on its surroundings. The influence must propagate in less than $[(\Delta t)_{dynamical} - t_{delay}]$, where $(\Delta t)_{dynamical}$ is the interval between the turnaround time of galactic mass perturbations of similar amplitudes in incipient clusters and in incipient voids. See text for further discussion.

$\delta_{cluster} > 0$) and the turnaround of an equivalent fraction in an incipient void (where $\delta_{void} < 0$) is $\Delta t/t \approx \delta_{cluster} + |\delta_{void}|$, where t is the cosmological timescale and both $\delta_{cluster}$ and $|\delta_{void}|$ are still < 1; unless the influence pervades the void within a time Δt it plainly cannot pre-empt the formation of galaxies there (see Figure 2). But it would seem even easier for a galaxy to quench information of new neighbours - whatever the actual feedback mechanism was it would surely be stronger at closer range, and the energy released within an incipient cluster might then be inadequate to exert any feedback on larger scales. This difficulty would not arise if

https://doi.org/10.1017/S0074180900136344 Published online by Cambridge University Press

there were a delay t_{delay} between the onset of protogalactic collapse (defined as the latest stage at which negative feedback could stop the protogalaxy from evolving into a luminous galaxy) and the stage when it would first exert negative feedback. The first galaxies would then be unable to quench the formation of others (even close neighbours) unless they lagged behind the first ones by longer than t_{delay}. Remote negative feedback would still occur if the influence of the first galaxies reached the void in a shorter time than $\Delta t - t_{delay}$. The physical interpretation of t_{delay} depends on the actual feedback mechanism. If, for instance, this involved quasar-like activity in the first galaxies, t_{delay} would be the time taken for a young galaxy to develop violent activity in its nucleus.

4. A NOTE ON THE LARGE-SCALE ENVIRONMENTAL EFFECTS OF QUASARS

The stellar IMF, and hence the M/L of a galaxy, may depend on whether or not the protogalactic medium is photoionized. High-redshift quasars are a much-discussed (and perhaps even dominant) contributor to the ionizing UV background. The large range of influence of a single long-lived quasar may then have interesting implications. A quasar with lifetime t_Q can ionize a spherical volume whose radius, expanded to the present epoch, is

$$R_I = 40\ h_{100}^{-2/3} \left(\frac{\Omega_b}{0.1}\right)^{-1/3} \left(\frac{L_{uv}}{10^{46}\mathrm{erg\ s^{-1}}}\right)^{1/3} \left(\frac{t_Q}{2\times10^9\mathrm{yrs}}\right)^{1/3} \mathrm{Mpc}$$

(if its radiation were beamed, a quasar with the same overall luminosity could ionize matter in a cone of even greater extent).

These dimensions exceed the correlation scale of gravitational clustering in the CDM model (which predicts that the correlation function for the mass distribution should go negative on scales exceeding 18 h_{100}^{-2} Mpc). If the efficiency of bright star formation were indeed modulated by the UV background, then quasars, through having photoionized the medium in a 'patchy' fashion, could conceivably have created large-scale correlations and imprinted apparent large-scale structure (bubbles, chains, etc.) of entirely non-gravitational origin, which would therefore induce no associated streaming velocities. (Similar effects could be induced at an earlier stage by the wakes of large-scale cosmic strings (Rees 1986b)).

5. CONCLUSIONS

Most biasing schemes involve dissipative gas dynamics and are therefore hard to quantify. Some kind of biasing is mandatory if we are to reconcile any flat $\Omega = 1$ model with the data on groups and clusters. There are many options: we cannot decide between them in our present ignorance about dark matter. We need, for instance, to know the distribution of galaxies of all types (including dwarf galaxies); what is in the voids; and whether there are galaxies without halos (or indeed, dark halos without luminous cores). To reconcile the CDM model with $\Omega = 1$ may not require anything beyond the "natural" biasing that arises from non-dissipative gravitational clustering. Whatever cosmogony turns out to be the right one, the notion that galaxies trace mass is almost certainly an unjustifiable assumption.

I am particularly grateful to Avishai Dekel for collaboration on topics touched on in this brief review, and to many colleagues, especially Carlos Frenk and Jim Peebles, for helpful discussions.

References

Bardeen, J.M., Bond, J.R., Kaiser, N. and Szalay, A. 1986 Astrophys. J. **304**, 15.
Bruzual, G. 1983 Astrophys. J. **273**, 105.
Dekel, A. and Rees, M.J. 1987 Nature **326**, 455.
Dekel, A. and Silk, J.I. 1986 Astrophys. J. **303**, 39.
Frenk, C.S., White, S.D.M., Efstathiou, G. and Davis, M. 1986 Nature **317**, 595.
Kaiser, N. 1984 Astrophys. J. (Lett.) **284**, L9.
Peebles, J. 1986 widely-circulated unpublished screed.
Rees, M.J. 1983 in 'Large Scale Structure of the Universe' ed. G. Abell and G. Chincarini p299 (Reidel, Dordrecht).
Rees, M.J. 1985 MNRAS **213**, 75P.
Rees, M.J. 1986a MNRAS **218**, 25P.
Rees, M.J. 1986b MNRAS **222**, 27P.
Silk, J.I. 1985 Astrophys. J. **297**, 1.
White, S.D.M., Davis, M., Efstathiou, G. and Frenk, C.S. 1987b Nature (in press).
White, S.D.M., Frenk, C.S., Davis M. and Efstathiou, G. 1987a Astrophys. J. **313**, 505.