ON THE DYNAMICAL EVOLUTION OF CLUSTERS OF GALAXIES

Jeremiah P. Ostriker
Princeton University Observatory
Princeton, New Jersey

1. INTRODUCTION: STELLAR SYSTEMS OF POINT MASSES

The theory of the dynamics of star clusters (cf. Spitzer 1975 for a review) is by now so well developed that we have, or think we have, a moderately accurate picture of the physical processes acting in and the overall evolution of spherical systems. In contrast, flattened and/or rotating systems are apparently subject to a variety of ill-understood instabilities which ultimately are a manifestation of the second law of thermodynamics; at given total energy, a system will tend to increase the fraction of its kinetic energy in disordered rather than ordered form. But spherical systems (globular clusters, elliptical galaxies, Morgan cD clusters of galaxies) are relatively smooth and featureless; they show little substructure indicating, presumably, that they are quite stable to perturbations of their fundamental normal modes, and they are normally modeled as rather "hot", pressure supported systems.

In fact, the confidence we have in our understanding of spherical systems has not been very well tested empirically. Not only do we have no direct information concerning dynamical evolution (less even than the indirect information available about evolution of stars from cluster H-R diagrams), but our information concerning equilibrium is quite poor. In the best observed systems we know a central velocity dispersion and the projection on the plane of the sky of the density distribution of a subsystem of tracer objects. In addition to the obvious loss of information due to projection and the unavailability of $[\langle v_{\parallel}(r)^2\rangle, \langle v_{\perp}(r)^2\rangle]$ in any system* we have growing evidence that the observable tracer population is unrepresentative (Spinrad et al. 1978), that its number distribution does not reflect the underlying mass distribution. The problem, however, of the evolution of a more or less spherical N-body system is essentially simple, it might, in principle, have been solved by Newton. And the well developed (if untested) theory for the dynamics of N point masses should provide a good starting

*In our galaxy, we do know that, locally $\langle v_{\parallel}{}^2\rangle/\langle v_{\perp}{}^2\rangle \sim 2$ for the Halo (spheroidal) population (Oort 1965).

M. S. Longair and J. Einasto (eds.), The Large Scale Structure of the Universe, 357–375.

point for the more complicated dynamics of N galaxies, each having many excitable internal degrees of freedom.

For the average star it is customary to distinguish between processes occurring on two widely separated time scales. On the dynamical time scale $\tau_D \approx (G\bar{\rho})^{-1/2}$, where ρ is the mass density, a typical star will cross the system, purely statistical fluctuations will change, and the system will respond to any gross departures from (virial) equilibrium. On the much longer relaxation time scale, $\tau_R \approx (N/\ell nN)\tau_D$, a star will interact with other individual stars (rather than the cluster as a whole). If external force fields (e.g., "tidal forces") significantly affect the orbits of test particles then there will be time scales associated with these as well.

Various stages in the evolution of an N-body system are customarily enumerated which occur on successively longer time scales. The first two listed are highly conjectural and other scenarios (cf. Doroshkevich et al. 1974) are just as plausible.

a) Uniform isotropic expansion.
Globular clusters, galaxies, cluster stars of galaxies are all thought to arise from perturbations growing in a standard Friedmann-Robertson-Walker cosmology.

b) Maximum size reached, star formation, collapse.
If a given perturbation has mean density greater than the cosmological critical density at a given epoch, $\rho_{crit} = (3H^2/8\pi G)$, then a maximum size will be reached, the perturbation will separate out from the general Hubble flow, and it will recollapse in a probably anisotropic fashion with further material falling into the deepening potential well (Gott and Gunn 1972). Much of the material may be processed into "point" objects (those of much higher than mean density) during these phases. That is, much star formation must occur during the collapse phase of galaxy or globular cluster formation (Eggen et al. 1962, Binney 1977, Peebles and Dicke 1968) and much galaxy formation must occur during or before the collapse phase of cluster formation (Rees and Ostriker 1977).

c) Violent relaxation.
The system of point masses so formed is generally out of equilibrium at first, and the violent large scale and probably asymmetrical motions which ensue produce strong gravitational fields, fluctuating in space and time. Individual particles interacting with these fields will clearly not preserve "integrals" of motion such as energy and angular momentum, and arguments can be given (Hénon 1964, Lynden-Bell 1967) for the establishment of quasi-Maxwellian distribution functions.

These first three phases occur on the dynamical time scale establishing a system of point masses more or less resembling the classical Emden isothermal sphere (cf. Zwicky 1957) but, of course, truncated in some way.

d) Equipartition, dynamical friction.
The (presumably few) relatively massive objects (mass M_h) in the central regions of the cluster interacting with the swarm of average particles of mass $\bar{M}$ (number N) suffer numerous gravitational collisions which lead towards equipartition. On the time scale of $(\bar{M}/M_h)\tau_R$ they fall towards the center of the system (cf. Spitzer 1969, Tremaine et al. 1975).

e) Core contraction.
Even in a system of particles all of the same mass there is a tendency for the central density to increase. The exact physical characterization of the process is still under debate (cf. Antonov 1962, Lynden-Bell and Wood 1968, Spitzer and Thuan 1972, Lightman and Shapiro 1977) but the results are agreed upon. On a time scale of $\sim 100\ \tau_R$ evaluated at the center, the central density approaches a singularity with the mass in the high density core approaching zero and the energy (kinetic or gravitational) of the core changing relatively slowly.

f) Then, on the considerably longer time scale of $\sim 100\ \tau_R$ evaluated at the half mass point (rather than the center), the cluster as a whole will change significantly, although it is not at all clear yet whether that change is better characterized as collapse or evaporation, simply because no treatments, analytical or numerical, have been able to reach past stage e).

2. PHYSICAL INTERACTIONS AMONG "SOFT" OBJECTS

All of the stages of evolution enumerated above would presumably occur in the great clusters of galaxies were there enough time. For Coma the dynamical time scale phases a) - c) is $\tau_D \approx 10^{9.3}$ yrs, the equipartition time scale d) for the most massive central galaxies $(\bar{M}/M_h)\tau_R \approx 10^{10.1}$ yrs and all the other processes are too slow to be of interest on the Hubble time scale. The dynamics of galaxies in galaxy clusters is, however, far more interesting than that of stars in star clusters. The reason for this is simply that the ratio of the size of the typical member to the distance of its nearest neighbor is $\sim 1/5$ for the galaxy in the center of a rich galaxy cluster but $< 1/10^5$ in even a very condensed globular cluster. Thus, gravitational collisions between galaxies can excite their internal degrees of freedom producing many effects not present in the point Newtonian dynamics of star clusters. In some respects the problems encountered are reminiscent of those in atomic physics in that analogs of "exchange collisions", "excited states", "pressure ionization", etc. must be considered. The "standard" processes enumerated in the previous section have, of course, been considered with reference to clusters of galaxies. For example, Peebles (1970) deals with (a)-(c) and White (1976) with (d). However, there are, in addition, three kinds of processes which are peculiar to, or far more important in galaxy clusters than in star clusters; and I shall concentrate here on those. It is possible to maintain the hypothesis that the apparent ordering of cluster properties into a sequence of types by Bautz and Morgan (1970) or Oemler (1974) is only a manifestation of the greater or lesser degree of dynamical evolution produced by these three processes: dynamical evolution gradually transforms Bautz-Morgan III clusters (like Virgo) to type II systems (like Coma) or type I systems (like Abell 2199).

a) Galaxy interactions with gas

It is well known that ellipticals and SO galaxies are more common in clusters than they are in the field, relatively more common in condensed clusters than in low density systems, and relatively more common in the central than the outer regions of rich clusters (cf. Melnick and Sargent 1977 for a recent discussion). Spitzer and Baade (1951) suggested, some time ago, that collisions between spiral galaxies could sweep out the gas, prevent further star formation and allow the spirals to gradually be transformed to SO systems. The discovery of thermal X-rays from galaxy clusters (Kellogg 1973) allowed variants of this process to be considered. Gunn and Gott (1972) noted that the ram pressure of the ambient gas would tend to sweep spirals clean of gas, an idea developed in more detailed calculations of Gisler (1976). Recently, Cowie (1977) has considered the thermal effects and showed that interactions with the very hot ambient gas will tend to heat and evaporate gas from galaxies in X-ray clusters.

Any and all of these processes appear capable of stripping gas from spirals and producing SO systems. If, further, Ostriker and Thuan (1975) are correct in their contention that the discs of spiral galaxies are to a significant extent secondary, produced by infall of gas processed through halo stars, then early and efficient gas removal by any of the above mechanisms would prevent formation of the discs at all, leaving only the spheroidal bulge components of the would-be spiral galaxies which, as many investigations have shown, are indistinguishable from ellipticals. Thus, the same process which increases the proportion of SO systems in rich clusters, will also increase the proportion of ellipticals. This chain of argument has been strongly reinforced by the finding (Serlemitsos et al. 1977) of X-ray iron emission lines from three rich clusters. The quantity of metal rich material seen in the clusters is, within the observational and theoretical uncertainties, just that which was ejected from the elliptical galaxies according to the models of Ostriker and Thuan (1975) and just that which, were it allowed to fall back into the elliptical galaxies as secondary discs, would give the latter the appearance of normal spiral or SO systems.

Stellar dynamical processes to transform a large fraction of spirals to SO or elliptical systems in clusters are not an attractive alternative. The principal reason is that the galaxies ultimately

https://doi.org/10.1017/S0074180900144821 Published online by Cambridge University Press

found in the central regions of great clusters cannot be distinguished from field galaxies in numerical simulations of galaxy clustering (cf. Aarseth, Gott and Turner 1977) until the galaxy is actually in a region of quite high velocity dispersion. In that environment, as many studies have shown, interactions tend to be weak and dispersive. High velocity collisions do not produce mergers and have little effect on the inner parts of galaxies; the main result of such encounters, as we shall describe in the next section, is to produce tidal stripping.

b) High velocity tidal interactions

Dynamically, it is useful to define "clusters" vs. "groups" by the velocity dispersions of the members, rather than by the (somewhat ambiguous) number counts, and we shall call a cluster of galaxies an assemblage within which the velocity dispersion among the members is substantially larger than the velocity dispersions of the stars within the constituent galaxies; for the Coma cluster, the two numbers are $V_{rms,cluster} \approx 1{,}000$ km/s and $V_{rms,galaxy} \approx 200$ km/s along the line of sight. Thus, since galaxies are "soft" and have maximum central potentials comparable to $V_{rms,galaxy}$, collisions between galaxies will generally by hyperbolic regardless of impact parameter, and deviations from rectilinear motion of the galaxy centers will be small. Gallagher and Ostriker (1972) and others have treated the problem in the impulsive approximation. Richstone (1975, 1976) analyzed the fast collisions problem with essentially the same scheme of approximations as Toomre and Toomre (1972) used for the slow, nearly parabolic, collisions appropriate for field galaxies. First the relative orbital motion of two rigid diffuse galaxies is computed. Then, within each galaxy, stars are sampled by a Monte Carlo technique and their orbits calculated during the galaxy-galaxy encounter as a restricted three-body problem. Then, by integrating over the orbital parameters in the final state and comparing with the comparable integrals in the pre-collision state, one can estimate the change induced in the galaxy by the collision. Specifically, since galaxies can be minimally characterized by a mass M, central density ρ_c and two radii, an inner core radius R_c and an outer tidal cutoff radius R_t, one can compare the changes (δM, $\delta\rho_c$, δR_c, δR_t) induced by a specific collision. Then one can integrate over the distribution of impact parameters, relative velocities and masses of the perturbing galaxies. Richstone found that the changes in the core properties were small and uncertain. The envelopes could be changed significantly. In general, the rate of change of any quantity Q(= M, ρ_c, etc.) can be represented approximately, but simply, by the following formula:

$$\frac{1}{Q}\frac{dQ}{dt} = [(n_p M_p)/M_t] R^2_{t,t} V_{cl,rms} C_Q \,, \qquad (1)$$

where (n_p, M_p) are the number density and mass of the perturbing galaxies (entering only as their total mass density) and $(M_t, R_{t,t})$ are the mass and tidal radius of the test galaxy being stripped. Note that this result is independent of $V_{rms,gal}$ here in the limit that

$(V_{gal,rms}/V_{cl,rms}) \ll 1$. The parameters C_Q represent the efficiency of the collisions and are, according to Richstone's Monte Carlo calculations,

$$C_M = -0.015 \pm 0.003 \quad ,$$
$$C_{Rt} = -0.014 \pm 0.006 \quad . \tag{2}$$

Richstone also finds surprisingly that $C_\rho > 0$, collisions tend to increase the central densities; the result is however quite uncertain. The rate at which the tidal radius is decreased can be rewritten simply as

$$\frac{dR_{t,t}}{dt} = -3.3 \times 10^{-3} \frac{\rho_{cl}}{\langle\rho_t\rangle} V_{cl,rms} \quad , \tag{3}$$

where $\langle\rho_t\rangle$ is the mean density of the galaxy within its tidal radius and ρ_{cl} is the local cluster density. Clearly the tidal radius will be reduced until $(T_{Hubble}/T_{dyn}) = 300\ (\langle\rho_t\rangle/\rho_{cl})(R_t/R_{cl})$. Such an effect may already have been detected in Strom and Strom's (1977) observations of the Coma cluster.

c) Accretion and Cannibalism

As noted earlier dynamical friction (Chandrasekhar 1943) will cause massive satellites to spiral into their parent galaxies (Ostriker and Tremaine 1975; Tremaine 1976) and in clusters the massive galaxies will tend to accumulate in the center (Ostriker and Tremaine 1975; White 1976, 1977; Ostriker and Hausman 1977). Both effects tend to increase the observed luminosity of galaxies. The latter process (the merging of galaxies in clusters) will tend to produce supergiant systems of low central surface brightness like the known cD systems. One can also show that the apparently non-statistical features seen at the bright end of the cluster luminosity function can be caused by cannibalism. What is the "normal" luminosity function?

Although galaxies of apparently normal character exist over an enormous range of brightness, there is, however, a characteristic luminosity L_* for galaxies, since the luminosity function $\phi(L)$ (such that the number of galaxies per unit volume having luminosity in the range $L \to L + dL$ is $\phi(L)dL$) for both field and cluster galaxies can be fit to a function of the form proposed by Schechter (1976):

$$\phi(L) = (n_*/L_*)(L/L_*)^{-\gamma}\exp(-L/L_*) \quad , \tag{4}$$

with the value of L_* varying little between galaxies in the field and those in dense clusters. We plan to show here how the characteristic L_* determines, not only the typical galaxy seen at the knee of the luminosity function, but also the first brightest galaxy in clusters which have undergone extensive dynamical evolution.

There are alleged to be characteristics of the observed luminosity function, at the bright end, which are inconsistent with any statistical model. The most ingenious study to date, by Tremaine and Richstone (1977), compared the expected values of $\sigma(M_1)$ and $\langle\Delta M_{12}\rangle \equiv (M_2 - M_1)$ in a way that did not depend on the explicit form of the luminosity function. Applying their analysis to the Sandage and Hardy (1973) cluster data, they tentatively agreed with Sandage's conclusion that the luminosity of the brightest cluster galaxies is determined by some special process. More explicitly, Tremaine and Richstone proved that for any statistical luminosity function, the inequality $\sigma(M_1)/\langle\Delta M_{12}\rangle \equiv t_1 > 1.0$ must hold. In fact, one can show that, for a Schechter luminosity function with $\gamma = +1.25$ and n_* in the interval 25 - 100, t_1 is fairly constant and equals 1.20 ± 0.02; in contrast, the value of t_1 derived from the Sandage-Hardy data, as presented in Tremaine and Richstone, is 0.48 ± 0.10 (standard error).

Three separate processes appear to be combining to produce this statistically quite significant effect.

1. The first brightest galaxies are somewhat better standard candles as measured through a fixed metric diaphragm than would be expected on the basis of the slope of the luminosity function at the bright end and the number of galaxies in the cluster (cf. Dressler 1977). The effect is most pronounced in centrally condensed spiral poor clusters (type I clusters).

2. The brightness of the first brightest galaxies depends much less on the cluster richness than would be expected with $(dM_1/d \log n_*) = -0.16 \pm 0.02$ (S.E.) observed, (derived from data presented in Richstone and Tremaine 1976) compared to $(dM_1/d \log n_*) = -0.9$ for the Schechter function with $5 < n_* < 50$ (cf. also Sandage 1976 and Schechter and Peebles 1976).

3. The gap between the first and second brightest galaxy is larger than would be expected statistically, it is larger than can be accounted for simply by saying the first galaxy is too bright, and it is also most pronounced in centrally concentrated type I clusters (cf. Dressler 1977). This last effect must be confirmed using more homogeneous data than has hitherto been available.

Sandage and Hardy (1973) commented that "The brighter the dominant galaxy becomes, the absolutely fainter will be the second and third ranked members. 'The rich are rich at the expense of the poor, progressively!'"

I would like to suggest here that that is exactly what is happening, that in type I clusters the cannibalism described by Ostriker and Tremaine (1975) and White (1976) will produce just these effects.

First let us treat the luminosity evolution of the first brightest galaxy by a rough qualitative argument. Consider a simplified case of the merging of N identical galaxies, each of mass M, core radius R(1)

https://doi.org/10.1017/S0074180900144821 Published online by Cambridge University Press

defined so that each has gravitational energy $W = -GM^2/R(1)$, and total energy $E = -GM^2/2R(1)$. If we combine N such galaxies, the total mass is NM and the total energy (neglecting the contribution produced by initial orbital binding energy) is $E(N) = -NGM^2/2R(1)$, which we may identify with $-G(NM)^2/2R(N)$ assuming that the new galaxy differs from its constituent parts only by scaling factors. This gives for the radius and mean surface brightness within $R, \Sigma(\propto(NM)/R^2)$

$$R(N) = NR(1) \; ; \quad \Sigma(N) = N^{-1}\,\Sigma(1) \quad ; \tag{5}$$

the luminosity and radius will increase, but the mean surface brightness will fall with increasing N (cf. Figs. 2a and 2b). Suppose now that we observe the growing galaxy through a diaphragm of radius R_D which is much larger than R(1). All the luminosity passes through the diaphragm and, as N increases, the observed luminosity will be proportional to N until $R(N) \approx R_D$. Thereafter the luminosity observed is $\approx \pi R_D^2 \Sigma$ and declines proportionally to N^{-1}, the peak apparent luminosity being

$$L_{max} \approx \pi R_D R(1)\Sigma(1) \approx \left(\frac{R_D}{R_s}\right) L_s \quad , \tag{6a}$$

where L_s and R_s are approximately the luminosity and characteristic radius of the standard building-block galaxies. We can make this argument more precise if we adopt a definite intensity profile for the component galaxies. The most accurate two-parameter fit is, according to Kormendy (1977), the de Vaucouleurs (1953) law which can be written $I(r) = 2141.5\ I_o \exp[-7.6692(r/R_o)^{1/4}]$, $L_{tot} = 22.666\ I_o R_o^{\,2}$, where (I_o, R_o) are the surface brightness and radius at the cylinder containing one-half of the total light. We can now compute the luminosity observed within R_D from the first brightest galaxy composed of N standard objects by integrating the intensity profile over the diaphragm area with $I_o = N^{-1} I_{os}$ (or $L_{1,tot} = NL_s$) and $R_o = NR_{os}$. The result is given by equation (12a). As N increases, $L_{1,obs}(N,R_D)$ reaches a broad flat maximum and then slowly declines for $N > N_c$. The maximum value is $L_{max,obs} = 0.7390\ (R_D/R_{os})L_s$ for the de Vaucouleurs model. Inclusion of the orbital (relative) energy affects the results only by scaling factors: we can estimate the magnitude of the effect by comparing this analytical treatment with the numerical results described later, and obtain

$$L_{max,obs} = 1.5\ (R_D/R_{o*})L_s \quad . \tag{6b}$$

For the de Vaucouleurs model the critical value N_c is

$$N_c = 6.084\ (R_D/R_{os}) \approx 14\ (R_D/R_{o*}) \tag{7}$$

In Fig. (1a) we show $L_{1,obs}$ as a function of $\ln(N/N_c)$ since the latter quantity is roughly proportional to the time (the accretion process being initially exponential); the analogous relation computed via the Monte Carlo simulation is displayed in Fig. (2c,d). We also show in (1a) the model $\ln(N/N_c)$ dependence of the important observable parameter

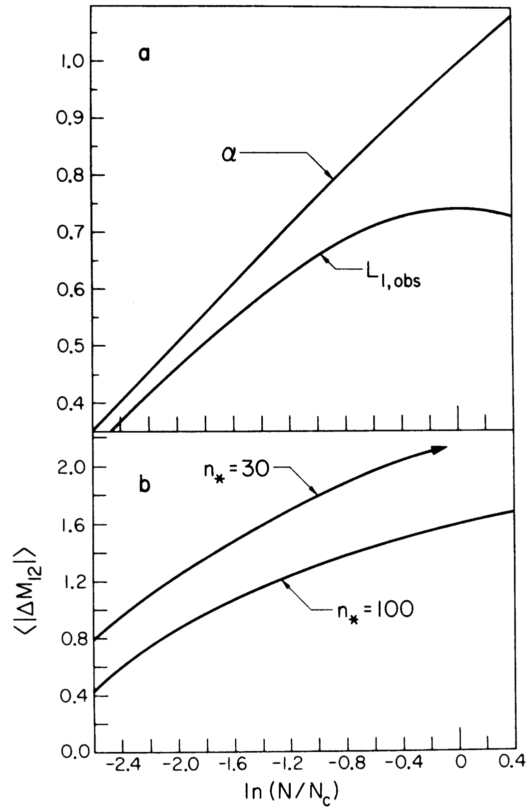

FIG. 1: (a) Observed luminosity of accreting galaxy plotted vs. number of accreted objects N according to theory of § IV B. The ordinate $L_{1,obs}/(L_s R_D/R_{os})$ is computed parametrically as $3459.5\ P(8,x)x^{-4}$ where $\ln(N/N_c) = 6.34282 - 4 \ln x$. Also plotted is $\alpha(R_D)$ defined by eq. (8) and computed as $x^8 e^{-x}/[20160\ P(8,x)]$. (b) The expectation value for the magnitude difference $\langle \Delta M_{12} \rangle$ for clusters of richness $n_* = 30,100$ computed by equations (7) and (10) - (13), assuming no luminosity evolution for any galaxy except the brightest.

$$\alpha(R_D) = \left(\frac{d\ln L_{obs}}{d\ln R}\right)_{R=R_D} , \qquad (8)$$

which is to be compared with Fig. (2e). Note that the scales in Fig. (1) are linear in luminosity (not magnitude) and logarithmic in (N/N_c); $L_{1,obs}$ is a very weak function of accretion once the latter is substantial. Note also that $\alpha(16)$, which for Oemler's sample ranges between 0.29 and 0.78, is monotonic and a good candidate for an observable to replace the unobservable "time" or (N/N_c) coordinate.

The value of the maximum luminosity is set partially by instrumental effects (R_D) and partially by the intrinsic properties of the standard galaxies contributing most to the growing supergiant. The maximum magnitude corresponding to the numerical value in equation (3) depends on R_{o*}. From the mean of 10 Coma cluster galaxies having $\bar{L} = L_*$ whose properties are measured in Oemler (1976), we find R_{o*} = 5.6 kpc. This would give a maximum luminosity within a 43 kpc diaphragm of $M_B = -23.2$. This is somewhat greater than the mean first brightest cluster galaxy for Bautz-Morgan I (highly evolved) clusters of -22.7 given by Sandage and Hardy (1973).

Gunn and Tinsley (1976) investigating the same problem also commented that evolution could in principle lead to either an increase or decrease in the apparent luminosity of the accreting galaxy. We see from the above calculation (see also Hausman and Ostriker 1977) that, for galaxies with properties like observed galaxies, $(dL_{obs}/dt) > 0$ simply because the effective core radii are always smaller than the radii of the diaphragms typically used by observers.

Next, let us look at the apparent luminosity evolution ΔM_{12}, the magnitude difference between the first and second brightest galaxies. The probability of cannibalism per unit time is proportional to the mass of the giant times the mass of the prospective victim (neglecting a slowly changing logarithmic factor). Thus the galaxy most likely to be eaten is the second brightest galaxy after which the new second brightest is the former third brightest and is consequently fainter; it follows that M_2 will become fainter and $|\Delta M_{12}|$ will increase with time. Again, a simple analysis allows us to be roughly quantitative. Let $\nu(\ell,t)dt$ be the probability that a galaxy with luminosity $\ell \equiv L/L_*$ is eaten in time dt. Ignoring the luminosity evolution of all except the first brightest galaxy (to be considered elsewhere), the probability of survival till time t is

$$p(\ell,t) = \exp(-\int_o^t \nu(\ell,t)dt) \quad , \qquad (9)$$

and the expectation value (ensemble average) of M_{12} is

$$<\Delta M_{12}> = p_2\Delta M_{12} + (1-p_2)p_3\Delta M_{13} + (1-p_2)(1-p_3)p_4\Delta M_{12}\cdots \qquad (10)$$

where all quantities are explicit functions of time and $p_k \equiv p(\ell_k,t)$ is the probability of survival of the cluster member which was originally kth rank in luminosity

$$\Delta M_{1k} = 2.5 \log\left(\frac{L_{1,obs}}{L_{k,obs}}\right) = 2.5\left[\log\left(\frac{L_{1,obs}}{L_s}\right)\left(\frac{\ell_s}{\ell_k}\right)\left(\frac{L_k}{L_{k,obs}}\right)\right] \qquad (11)$$

$$\frac{L_{1,obs}}{L_s} = NP[8, 7.6692\ (R_D/NR_{os})^{1/4}]/7! \quad , \qquad (12a)$$

https://doi.org/10.1017/S0074180900144821 Published online by Cambridge University Press

$$\frac{L_{k,obs}}{L_k} = P[8,7.6692\ (R_D/R_{ok})^{1/4}]/7!\ ,\ \frac{L_s}{L_*} = (2 + \delta - \alpha) \qquad (12b)$$

where P(x,y) is the incomplete gamma function. Here δ is the exponent in the assumed mass luminosity relation $(M/L) = (M_*/L_*)\ell^{\delta}$ and the most probable values of ℓ_k can be obtained from the formulae in Schechter (1976). The total rate of accretion by the central object is given by

$$dN(t) = n_* \nu_D(\ell_*,t)dt \int_0^{\infty} \ell^{-\alpha+1+\delta} e^{-\ell} d\ell = \qquad (13)$$

$$-n_* \ell^{-(1+\delta)} \Gamma(2+\delta-\alpha) d\ln p(\ell,t) \quad ,$$

giving $p_k = \exp[-(N\ell_k{}^{1+\delta})/n_*\Gamma(2+\delta-\alpha)]$ from equation (7) since $\nu(\ell,t) = \ell^{1+\delta}\nu(\ell_*,t)$. The uncertain time variable has fortunately been bypassed; given equations (7) and (10)-(13) we can evaluate $\langle\Delta M_{12}\rangle$ as a function of (N/N_c) or α for given values of (R_D,R_{os}) and $n*$. Taking (43 kpc, 5.6 kpc) for the former we show the expected evolution of $\langle\Delta M_{12}\rangle$ in Fig. (1b) for two values of n_*, these curves give upper bounds on $|\Delta M_{12}|$, since luminosity evolution of $L_2(L_3$, etc.) has been ignored. A direct check of the proposed correlation between ΔM_{12} and α seen in Fig. (1) is possible and would test the theory presented here.

Thus we expect that, as dynamical evolution proceeds, the first brightest galaxy will initially gorw in core radius and luminosity; its core radius will continue to grow, but the luminosity seen through a metric diaphragm will level out near L_{max} (equation 6) and then gradually decline; the total luminosity will steadily increase. The other bright galaxies in the cluster core will tend to be swallowed and the gap between the first and second brightest galaxy will grow steadily (equation 10).

Finally, let us compare different clusters at the same epoch. For fixed surface brightness, or distinctiveness compared to the background, the relaxation time T_E, which varies as $v_{rms}^3/(G^2 M_{gal}\rho_{cl})$, tends to decrease with decreasing richness as $N^{5/4}$. Thus there is relatively more dynamical evolution in poor clusters than in rich (at the same surface brightness) which consequently increases $(-M_1)$ for the poor clusters and reduces the expected amount of the Scott effect. The Bautz-Morgan classification (1970), which essentially measures cluster dynamical evolution, should be correlated with T_E (and crossing time, T_{Dyn}). It would be interesting to test the proposed correlation for those clusters having two of the three dynamical parameters (R_{core}, N_{core}, v_{rms}) measured as well as Bautz-Morgan type or ΔM_{12}.

Hausman and Ostriker (1977) have numerically simulated cluster evolution. They assume that 1) All galaxies have a surface brightness obeying the Hubble law $I(r) = I_0(1 + r/\beta)^{-2}$ with mass density proportional to light emissivity, the galaxy being truncated where its density reaches the cluster density. We define a core luminosity $L_0 \equiv I_0\beta^2$, and assume that $(M/L_0) \propto L_0^{0.5}$ (cf. Faber and Jackson 1976) and $\beta \propto L_0^{0.1}$ (cf. Oemler 1976). The relation between (U-B), (B-V) and M_V is taken from Sandage (1972). 2) The initial distribution of galaxy luminosity L is given by equation (1) with $\gamma = +1.0$ and with $L_* = 3 \times 10^{10} L_0$. 3) The accretion rate is given by Ostriker and Tremaine (1975) formulae which, under these circumstances, give $\dot{M}_1 \propto M_1 \int M_{c1} dn_{c1}$ where M_1 is the growing central giant and n_{c1} and M_{c1} are the number and mass of cluster galaxies of various types. 4) Collisions conserve mass, energy (binding + orbital), and luminosity in the (U,B,U) bands. We take an initial assumedly central first brightest galaxy with $L = 3L_*$ then, in a given time step, pick a victim galaxy from the Schechter distribution, the probability being proportional to the mass of the galaxy, find the new, swollen, first brightest galaxy from the conservation laws noted above, and repeat the process. Accretion is intrinsically unstable; if the first galaxy eaten happens to be particularly large, the primary will subsequently eat at a more rapid rate, and vice versa. Thus a Monte Carlo stimulation is useful. A detailed description of the numerical procedures and more extensive publication of results is reserved for a subsequent paper (Hausman and Ostriker 1977).

Figures (2a)-(2e) show the evolution of the first brightest galaxy as observed through diaphragms of radii 16 kpc and 30 kpc (for comparison, Gunn and Oke 1975; Sandage and Hardy 1975, use 19 kpc and 43 kpc for a Hubble constant of 50 km^{-1} Mpc^{-1}). Five Monte Carlo simulations are shown for identical starting conditions. A major uncertainty is how to treat the accreted galaxies when their mean density is more than the central density of the growing cannibal, which decreases continuously. Accretion will still occur and the initial relative orbital energy is still available, but tidal forces will not necessarily disrupt the accreted galaxies (unless they disrupt each other) so the self-binding energy may not be available and parts of the accreted galaxies may remain as intact cores to be seen as the "multiple nuclei" often found in cD systems. To simplify the discussion here, we terminated the integrations displayed when these effects became significant.

In Fig. (2c) the leveling off of the observed luminosity discussed in §2 is seen clearly; the five runs give a mean L_{max} of [4.78 ± 0.23 (standard error)] L_* compared to 2.2 L_* derived from equation (3), the differencebeing due to the neglect of the orbital binding energy in the analytical calculation. Figure (2a) illustrates the exponential growth of the total luminosity ($N \propto e^t$), which is to be compared with observed isophotal luminosities, and (2e) the approximately linear behavior of α with t (or lnN). In Figures (3a) and (3b) we show the expected luminosity evolution as a function of α (equation 8). Notice that the dispersion is smaller since α effectively measures dynamical evolution.

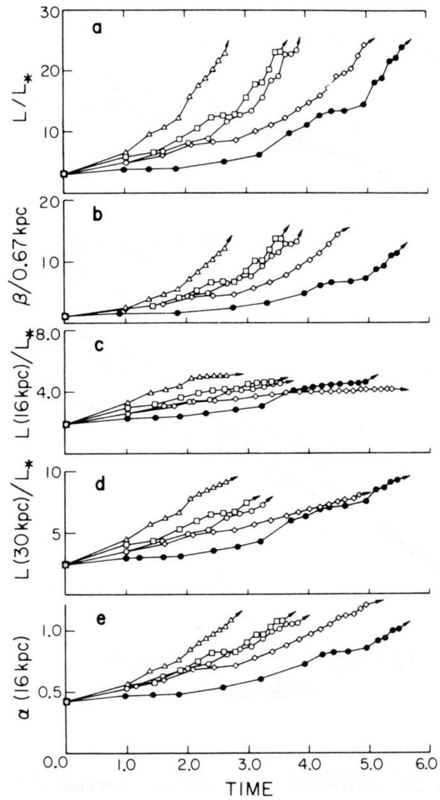

FIG. 2: Five Monte Carlo runs for evolution of first brightest cluster galaxy with initial luminosity $L = 3L_*$ (see § IV C for details). (a) Total luminosity; each successive symbol represents an accretion of a single galaxy. (b) Core Hubble radius vs. time; galaxy with $L = L_*$ has $\beta = 0.67$ kpc initially. (c), (d) Luminosities observed through 16 kpc and 30 kpc diaphragms. (e) $\alpha(R_D)$, the dimensionless measure of the core radius (eq. 8).

Hubble radii are given in Oemler (1976) for 6 galaxies which are first brightest in their clusters (MKW4, Virgo, A779, A1413, A2147 and A2670). From these, estimates for (α, L) can be obtained which are accurate to the extent that the galaxies fit Hubble laws in the range of radii considered. For this sample the mean and dispersions in $(\alpha(16), L(16)(\alpha(30), L(30))$ were calculated and are also shown in Fig. (3). The agreement is excellent, but of course, may be fortuitous.

Figure (4) shows core radii vs. luminosity as derived from the Monte Carlo simulations plotted with Oemler's data. The large increase in core radii observed for central luminous galaxies follows naturally from the dynamical theory. Figure 5 shows the expected evolution of

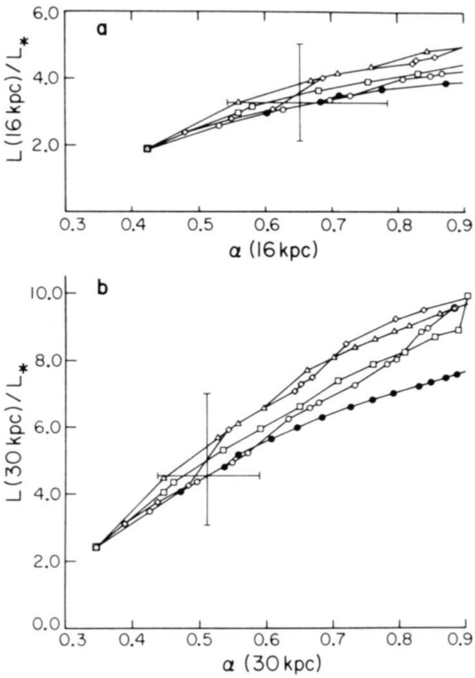

FIG. 3: Observed luminosities plotted versus α from Monte Carlo runs. Large cross data point is the mean of 6 first brightest galaxies from Oemler (1976).

the (U - B) vs. M_V relation. The bright galaxies do get bluer with time, but at only a moderate rate since they tend to preferentially accrete rather luminous ($L \approx 1.25\ L_*$) red galaxies.

In a further numerical investigation (Hausman and Ostriker 1977) finite clusters with various values of n_* were simulated by a Monte Carlo process and then allowed to interact according to the previously described rules. Here it was found that the luminosity growth of the first brightest galaxy tended to limit out at $|M_1 - M_*| = 1.5 - 2.2$ mag, for $n_* = 30$-100, rather than to continue to increase exponentially, the reason of course being that the supergiant ultimately "uses up" all of the easily available victims. The luminosity function change is quite interesting. It tends to steepen at the bright end and to develop an extra peak of super bright (cD) systems bearing some resemblance to the results found by Oemler (1976) and Dressler (1976).

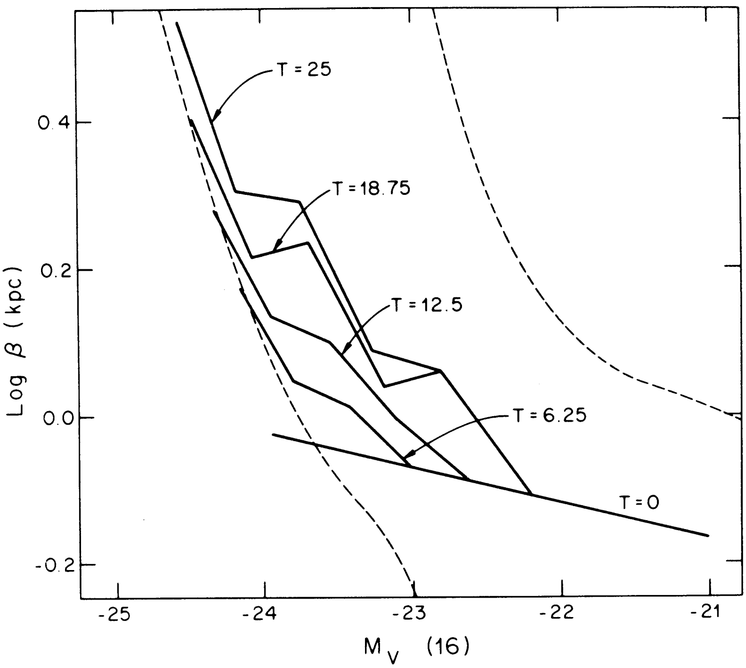

FIG. 4: Core Hubble radius versus magnitude from Monte Carlo simulations (T is time coordinate): data from Oemler (1976) lie between dashed lines with means near T = 0, T = 18.75 lines.

In sum, it appears that many of the notable features of centrally condensed clusters of galaxies, particularly the presence of very luminous but low surface brightness central cD systems, having a relatively small luminosity dispersion through fixed metric diaphragms and an apparently too large gap between first and second or third brightest galaxies, can be understood in terms of a straightforward dynamical theory of galactic cannibalism.

Bright galaxies spiral to the center of clusters on an equipartition time scale; there they are swallowed by a central giant which becomes physically bigger and brighter. Many tests of the theory are possible essentially by checking on correlations between α, ΔM_{12}, and relaxation or crossing times among observed clusters. In addition, since cannibalism tends, selectively, to deplete the bright end of the luminosity function, we can predict that the latter will become steeper and the break ($L*$) displaced to lower luminosity along the Bautz-Morgan sequence from Types III to I. To the extent that the theory is confirmed, it should be possible to use it, with accumulated

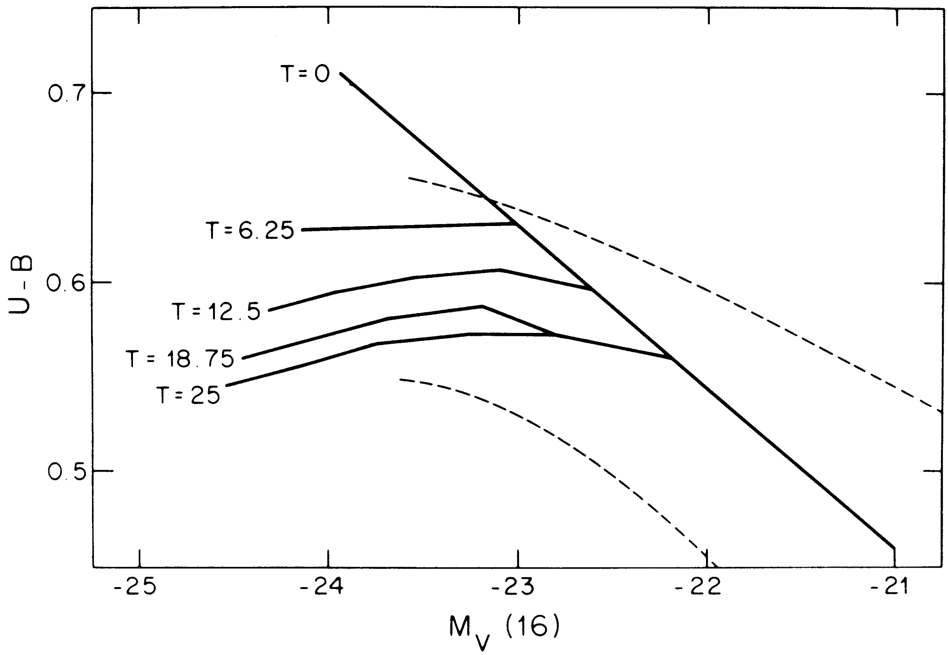

FIG. 5: Evolution of the color-magnitude relation from Hausman and Ostriker (1977). Dashed lines represent, approximately, the envelope from Sandage's (1972) observations.

data, to correct for the effects of dynamical evolution, thereby removing that source of dispersion and confusion from the Hubble diagram used in cosmological investigations. Since the Bautz-Morgan (1970) classification is correlated with ΔM_{12} (Dressler 1977), and ΔM_{12} is a good measure of dynamical evolution (for fixed n_*), the Bautz-Morgan corrections currently being used by Sandage and others already compensate for dynamical evolution, to some degree, on an empirical basis.

3. SUMMARY

The presently observed properties of clusters of galaxies have recently been reviewed by Bahcall(1977). These observations, and the theory of the last section, might be put together to sketch out a hypothetical evolution to this state in the form of the following tale:

Once upon a time, at the relatively recent epoch of $3 < z < 10$, galaxies formed from fluctuations of unknown origin in the expanding universe. Typical galaxies were very large ($10^{5.5}$ kpc) and massive ($10^{12.3}$ $M_\odot$). Large scale fluctuations took longer to develop, and clusters of galaxies separated out at the very recent epoch of $1 < z < 3$. The inner parts of the clusters collapsed, violently relaxed and adjusted to a

https://doi.org/10.1017/S0074180900144821 Published online by Cambridge University Press

nearly isothermal state. The outer parts continue to expand but at a decelerating rate. In the inner parts interactions with cluster gas (largely expelled from other galaxies) strips gas from galaxies and inhibits the formation of secondary discs (spirals and SOs). Relatively fast tidal interactions between galaxies strips off the dark material from the outer parts of individual galaxies leaving systems with conventional sizes and masses ($10^{4.5}$ pc, 10^{11} M_0) and distributing the dark matter throughout the inner parts of the cluster. Then on a longer time scale the giant systems tend to accumulate at the center as supergiant low surface brightness cD systems. The type of a given cluster in, for example, the Bautz-Morgan system is determined by the degree to which these dynamical processes have acted. Some depend more closely on the crossing time in a system, some on the relaxation time but in general the most evolved clusters should be the dense but poor (small N) systems. In extreme cases these dynamical processes may go to completion with the result that the whole cluster is transformed into one supergiant cD system containing remnant cores of destroyed galaxies.

I would like to thank Drs. A. Dressler, J. E. Gunn, A. Sandage L. Spitzer and S. Tremaine for useful conversations. I benefited from a John Sherman Fairchild Fellowship at California Institute of Technology, and the work was supported also by National Science Foundation grant AST76-20255.

REFERENCES

Aarseth, S., Gott, R., and Turner, E.: 1977, in preparation.
Antonov, V. A.: 1962, Vestn. Leningr. Gros. Univ. **7**, 135.
Bahcall, N.: 1977, Ann. Rev. Astron. & Astrophys. **15**
Bautz, L. and Morgan, W. W.: 1970, Astrophys. J. Letters **162**, L149.
Binney, J.: 1977, Monthly Notices Roy. Astron. Soc. in press.
Chandrasekhar, S.: 1943, Astrophys. J. **97**, 251.
Cowie, L.: 1977, Nature **266**, 501.
Doroshkevich, A. G., Sunyaev, R. A., and Zeldovich, Y. B.: 1974, Confrontation of Cosmological Theories with Observational Data, ed. M. S. Longair, IAU Symposium 63 (Reidel: Dordrecht).
Dressler, A.: 1976, Thesis, Lick Observatory.
Dressler, A.: 1977, preprint.
Eggen, O. J., Lynden-Bell, D., and Sandage, A. R.: 1962, Astrophys. J. **136**, 748.
Gallagher, J. and Ostriker, J. P.: 1972, Astron. J. **77**, 288.
Gisler, G. R.: 1976, Astron. & Astrophys. **51**, 137.
Gunn, J. E. and Gott, J. R.: 1972, Astrophys. J. **176**, 1.
Gunn, J. E. and Oke, J. B.: Astrophys. J. **195**, 255.
Gunn, J. E. and Tinsley, B.: 1976, Astrophys. J. **210**, 1.
Hausman, M. A. and Ostriker, J. P.: 1977, in preparation.
Henon, M.: 1964, Ann. Astrophys. **27**, 83.
Kellogg, E. M.: 1973, IAU Symposium #55.
Kormendy, J.: 1977, preprint.
Lightman, A. and Shapiro, S.: 1977, preprint.
Lynden-Bell, D.: 1967, Monthly Notices Roy. Astron. Soc. **136**, 101.

Lynden-Bell, D. and Wood, R.: 1968, Monthly Notices Roy. Astron. Soc. 138, 495.
Melnick, J. and Sargent, W. L. W.: 1977, Astrophys. J. in press.
Oemler, A.: 1974, Astrophys. J. 194, 1.
Oemler, A.: 1976, Astrophys. J. 209, 693.
Oort, J.: 1965, Stars and Stellar Systems Vol. V, ed. A. Blaauw and M. Schmidt (Chicago: U of Chicago Press), Chapt. 21.
Ostriker, J. P. and Hausman, M.: 1977, Astrophys. J. Letters, in press.
Ostriker, J. P. and Thuan, T. X.: 1975, Astrophys. J. 202, 353.
Ostriker, J. P. and Tremaine, S.: 1975, Astrophys. J. Letters 202, L113.
Peebles, P. J. E.: 1970, Astron. J. 75, 13.
Peebles, P. J. E. and Dicke, R. H.: 1968, Astrophys. J. 154, 891.
Rees, M. J. and Ostriker, J. P.: 1977, Monthly Notices Roy. Astron. Soc. 174, 520.
Richstone, D. O.: 1975, Astrophys. J. 200, 535.
Richstone, D. O.: 1976, Astrophys. J. 204, 642.
Sandage, A.: 1972, Astrophys. J. 176, 21.
Sandage, A. and Hardy, E.: 1976, Astrophys. J. 183, 743.
Schechter, P.: 1976, Astrophys. J. 203, 297.
Schechter, P. and Peebles, P. J. E.: 1976, Astrophys. J. 209, 670.
Serlemitsos, P. J., Smith, B. W., Boldt, E. A., Holt, S. S., and Swank, J. H.: 1977, Astrophys. J. 211, L63.
Spinrad, H. and Ostriker, J. P.: 1978, in preparation.
Spitzer, L.: 1969, Astrophys. J. Letters 158, L139.
Spitzer, L.: 1975, IAU Symposium #69, Dynamics of Stellar Systems, ed. A. Hayli (Dordrecht: Reidel), 3.
Spitzer, L. and Baade, W.: 1951, Astrophys. J. 113, 413.
Spitzer, L. and Thuan, T. X.: 1972, Astrophys. J. 175, 37.
Strom, S. and Strom, K.: 1977, Yale Univ. Symposium on Galactic Evolution.
Toomre, A. and Toomre, J.: 1972, Astrophys. J. 178, 623.
Tremaine, S.: 1976, Astrophys. J. 203, 72.
Tremaine, S., Ostriker, J. P., and Spizer, L.: 1975, Astrophys. J. 196, 407.
Tremaine, S. and Richstone, D.: 1977, Astrophys. J. 212, 311.
de Vaucouleurs, G.: 1953, Monthly Notices Roy. Astron. Soc. 113, 134.
White, S.: 1976, Monthly Notices Roy. Astron. Soc. 174, 19.
White, S.: 1977, Monthly Notices Roy. Astron. Soc. 179, 33.

DISCUSSION

Silk: Gas accretion by galaxies in rich clusters followed by subsequent star formation also can result in the luminosity increase of dominant cluster galaxies with time. The resulting effect on galaxy colours, however, may be different from that predicted in the cannibalism picture. For example, while gas accretion may initially result in a substantial "bluening" of the colours, this will be pronounced at redshifts ~ 1. If star formation ceased more than 3×10^9 years ago, the effect of gas accretion will be to produce net "reddening" (since intracluster gas is found to be enriched and is inferred to have originated early in the evolution of the cluster). My impression of the most recent data is that galaxy colours continue to redden with increasing luminosity. How do these data compare with the predicted flattening of the colour-magnitude relation that results from galaxy mergers?

Ostriker: This has often been raised as an objection to the theory and I would like to show what one would expect. The observations show that the brighter galaxies are redder and hence if massive galaxies grow by swallowing less massive galaxies, the bright galaxies should become bluer. However, since most of the galaxies being swallowed have luminosities about L*, the massive galaxies should evolve at roughly constant colour. If you see blue massive galaxies, they must be due to recent star formation.

Chernin: Clusters are in equilibrium, but only dynamically, i.e. they are gravitationally bound, not in statistical equilibrium. Relaxation via violent interactions may be rapid enough, but it cannot lead to the mass segregation observed in clusters.

Ostriker: I agree. The cluster centres are in dynamical equilibrium and only for the most massive galaxies in some clusters has equipartion begun to occur.

Tinsley: Ostriker has discussed how cannibalism of smaller elliptical galaxies by central cluster members will cause the latter to be somewhat too blue for their luminosity; this effect will lead to a levelling-off, at the bright end, of the colour-magnitude relation for E galaxies. I wish to mention that the only theoretical explanation of the colour-magnitude relation (Larson 1974, *Mon. Not. R. astr. Soc.*) predicts an intrinsic flattening of the relation at the bright end. The point is that smaller E galaxies are bluer because they lose their gas by supernova-driven winds at an earlier stage of chemical enrichment; the effect is insensitive to galaxy mass at the high-mass end because big galaxies retain their gas for long enough to make the mean stellar metallicity essentially the yield value (independently of the small gas fraction lost). Therefore, any observed levelling-off of the colour-magnitude relation for cluster galaxies may not be entirely due to cannibalism.

Ostriker: I agree.

Ozernoy: Could you talk a little more about the cD galaxies in poor clusters and especially those in the field? What about their origin? Should their properties differ appreciably from cD galaxies in rich clusters?

Ostriker: Yes. They should be less luminous and have smaller internal velocity dispersions.

Lynden-Bell: In the small clusters should the cD galaxies not be bluer in your theory?

Ostriker: Yes.

https://doi.org/10.1017/S0074180900144821 Published online by Cambridge University Press