

## Central Lancashire Online Knowledge (CLoK)

Title	Galaxy And Mass Assembly (GAMA): the life and times of L* galaxies
Type	Article
URL	<a href="https://clock.uclan.ac.uk/10079/">https://clock.uclan.ac.uk/10079/</a>
DOI	<a href="https://doi.org/10.1093/mnras/stt156">https://doi.org/10.1093/mnras/stt156</a>
Date	2013
Citation	Robotham, A. S. G., Liske, J., Driver, S. P., Sansom, A. E., Baldry, I. K., Bauer, A. E., Bland-Hawthorn, J., Brough, S., Brown, M. J. I. et al (2013) Galaxy And Mass Assembly (GAMA): the life and times of L* galaxies. Monthly Notices of the Royal Astronomical Society, 431 (1). pp. 167-193. ISSN 0035-8711
Creators	Robotham, A. S. G., Liske, J., Driver, S. P., Sansom, A. E., Baldry, I. K., Bauer, A. E., Bland-Hawthorn, J., Brough, S., Brown, M. J. I., Colless, M., Christodoulou, L., Drinkwater, M. J., Grootes, M. W., Hopkins, A. M., Kelvin, L. S., Norberg, P., Loveday, J., Phillipps, S., Sharp, R., Taylor, E. N. and Tuffs, R. J.

It is advisable to refer to the publisher's version if you intend to cite from the work.  
<https://doi.org/10.1093/mnras/stt156>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLoK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>

# Galaxy And Mass Assembly (GAMA): the life and times of $L^*$ galaxies

A. S. G. Robotham,<sup>1,2\*</sup> J. Liske,<sup>3</sup> S. P. Driver,<sup>1,2</sup> A. E. Sansom,<sup>4</sup> I. K. Baldry,<sup>5</sup>  
A. E. Bauer,<sup>6</sup> J. Bland-Hawthorn,<sup>7</sup> S. Brough,<sup>6</sup> M. J. I. Brown,<sup>8</sup> M. Colless,<sup>6</sup>  
L. Christodoulou,<sup>9</sup> M. J. Drinkwater,<sup>10</sup> M. W. Grootes,<sup>11</sup> A. M. Hopkins,<sup>6</sup>  
L. S. Kelvin,<sup>1,2,12</sup> P. Norberg,<sup>13</sup> J. Loveday,<sup>9</sup> S. Phillipps,<sup>14</sup> R. Sharp,<sup>15</sup>  
E. N. Taylor<sup>7,16</sup> and R. J. Tuffs<sup>11</sup>

<sup>1</sup>*SUPA†, School of Physics & Astronomy, University of St Andrews, North Haugh, St Andrews KY16 9SS, UK*

<sup>2</sup>*ICRAR‡, The University of Western Australia, 35 Stirling Highway, Crawley, WA 6009, Australia*

<sup>3</sup>*European Southern Observatory, Karl-Schwarzschild-Str. 2, D-85748 Garching, Germany*

<sup>4</sup>*Jeremiah Horrocks Institute, University of Central Lancashire, Preston PR1 2HE, UK*

<sup>5</sup>*Astrophysics Research Institute, Liverpool John Moores University, Egerton Wharf, Birkenhead CH41 1LD, UK*

<sup>6</sup>*Australian Astronomical Observatory, PO Box 296, Epping, NSW 1710, Australia*

<sup>7</sup>*Sydney Institute for Astronomy, School of Physics, University of Sydney, NSW 2006, Australia*

<sup>8</sup>*School of Physics, Monash University, Clayton, Victoria 3800, Australia*

<sup>9</sup>*Astronomy Centre, University of Sussex, Falmer, Brighton BN1 9QH, UK*

<sup>10</sup>*Department of Physics, The University of Queensland, Brisbane, QLD 4072, Australia*

<sup>11</sup>*Max Planck Institute for Nuclear Physics (MPIK), Saupfercheckweg 1, D-69117 Heidelberg, Germany*

<sup>12</sup>*Institut für Astro- und Teilchenphysik, Universität Innsbruck, Technikerstraße 25, A-6020 Innsbruck, Austria*

<sup>13</sup>*Institute for Computational Cosmology, Department of Physics, Durham University, South Road, Durham DH1 3LE, UK*

<sup>14</sup>*HH Wills Physics Laboratory, University of Bristol, Tyndall Avenue, Bristol BS8 1TL, UK*

<sup>15</sup>*Research School of Astronomy & Astrophysics, The Australian National University, Cotter Road, Weston Creek, ACT 2611, Australia*

<sup>16</sup>*School of Physics, The University of Melbourne, Parkville, VIC 3010, Australia*

Accepted 2013 January 23. Received 2013 January 8; in original form 2012 November 26

## ABSTRACT

In this work, we investigate in detail the effects the local environment (groups and pairs) has on galaxies with stellar mass similar to the Milky Way ( $L^*$  galaxies). A volume limited sample of 6150 galaxies are visually classified to determine the emission features, morphological type and presence of a disc. This large sample allows for the significant characteristics of galaxies to be isolated (e.g. stellar mass and group halo mass), and their codependencies determined.

We observe that galaxy–galaxy interactions play the most important role in shaping the evolution within a group halo; the main role of halo mass is in gathering the galaxies together to encourage such interactions. Dominant pair galaxies find their overall star formation enhanced when the pair’s mass ratio is close to 1; otherwise, we observe the same galaxies as we would in an unpaired system. The minor galaxy in a pair is greatly affected by its companion galaxy, and while the star-forming fraction is always suppressed relative to equivalent stellar mass unpaired galaxies, it becomes lower still when the mass ratio of a pair system increases.

We find that, in general, the close galaxy–galaxy interaction rate drops as a function of halo mass for a given amount of stellar mass. We find evidence of a local peak of interactions for Milky Way stellar mass galaxies in Milky Way halo mass groups. Low-mass haloes, and in particular Local Group mass haloes, are an important environment for understanding the typical evolutionary path of a unit of stellar mass.

We find compelling evidence for galaxy conformity in both groups and pairs, where morphological type conformity is dominant in groups, and emission class conformity is dominant in pairs. This suggests that group scale conformity is the result of many galaxy encounters

\* E-mail: asgr@st-andrews.ac.uk

† Scottish Universities Physics Alliance.

‡ International Centre for Radio Astronomy Research.

over an extended period of time, while pair conformity is a fairly instantaneous response to a transitory interaction.

**Key words:** large-scale structure of Universe.

## 1 INTRODUCTION

$L^*$  galaxies exist at the knee of the popular Schechter (1976) parametrization of the galaxy luminosity function. The galaxy stellar mass function has a similar form, and as such the knee of the stellar mass function ( $\mathcal{M}^*$ ) represents broadly the same galaxy population. Due to the shape of the galaxy luminosity function,  $L^*$  galaxies dominate the integrated mass and luminosity densities in the Universe at low redshift (e.g. Baldry et al. 2012; Driver et al. 2012).

To understand the integrated properties of all galaxies we can, to the first order, just consider the properties that affect the evolution of  $\sim L^*$  galaxies. Simply doing this offers insight into the fate of the median and modal unit of stellar mass in a cosmological volume. Further to this, apparent magnitude limited surveys at moderate redshifts (e.g. 2dFGRS, SDSS, MGC, GAMA, see Colless et al. 2001; Liske et al. 2003; Adelman-McCarthy et al. 2008; Driver et al. 2011, respectively) are dominated by  $L^*$  galaxies in terms of raw number counts, even though they are not the most populous type of galaxy. Thus, they are both extremely important in determining integrated properties of mass and light in cosmological volumes of the Universe, and the most easily studied variety of galaxy by virtue of the double-power-law shape of the luminosity function (see Cooray & Milosavljević 2005, for a discussion on the origin of  $L^*$ ). By a happy coincidence, our own galaxy, the Milky Way (MW), is also an  $L^*$  galaxy. This observational bias is not surprising, since it just means our Sun resides in the most typical type of galaxy to find stellar mass.

The combination of these factors makes a detailed investigation of  $L^*$  galaxies an instructive exercise, informing us both about the integrated properties of galaxies in cosmological volumes and about the typicality of our own galaxy. It is important that we fully understand any statistical biases present in the MW since, by virtue of proximity, it will always be the galaxy we can investigate in the most detail, and from which we can derive the most information about its formation history.

However, question marks remain over how typical the MW halo is in the context of the Universe and how unusual its galaxy occupation statistics are (e.g. Boylan-Kolchin et al. 2009, 2010; Boylan-Kolchin, Bullock & Kaplinghat 2011; James & Ivory 2011; Tollerud et al. 2011; Weisz et al. 2011; Lovell et al. 2012; Robotham et al. 2012). We are set to learn a vast amount about the MW in the coming decades. In the near future, *GAIA* (Wilkinson et al. 2005) will measure space motions and properties for two billion stars in the Local Group (LG) which includes all known member galaxies as part of the new era of Galactic archeology studies. Among likely discoveries, we will learn about dynamical equilibrium, or lack of it, for the first time. Building up to these hugely detailed surveys it is important we discover where the MW fits into the bigger picture. Only then can we apply what we know about the MW to cosmological galaxy formation models. Combining near-field cosmology (LG scale) and far-field cosmology (redshift surveys) is key to completing the full picture of galaxy formation (Freeman & Bland-Hawthorn 2002).

$L^*$  galaxies are observed to have extremely wide ranging properties, inhabiting a regime where we see the most variation in star formation, morphology and colour. This indicates that there is no

typical evolutionary history for these ‘typical’ galaxies. We can theorise about a large number of factors that might directly alter the evolution of any galaxy, including  $L^*$  ones. A non-comprehensive list includes galaxy–galaxy interactions, local scale environment (e.g. its dark matter halo), and the large scale environment of the cosmic web.

Many recent studies have investigated the role of environment in driving the evolution of galaxies, with much contradiction in the literature. Studies on SDSS by Weinmann et al. (2006) suggest strong evidence of ‘conformity’ between central and satellite galaxies in haloes for a given amount of stellar mass, the suggestion from Wang & White (2012) being that redder central galaxies of a given stellar mass tend to occupy larger haloes. This conclusion suggests that the halo environment has a significant role in determining the properties of galaxies, on top of those driven purely by stellar mass. Counter to this is recent work by Peng et al. (2010) using z-COSMOS and Thomas et al. (2010), who both broadly conclude that once ‘environment’ has influenced the stellar mass of a galaxy (particularly a central galaxy), the remaining properties are predominantly driven by the stellar mass. This is in alignment with earlier work by Baldry et al. (2006), which finds the colour–stellar mass and colour–concentration locus of galaxies is uncorrelated with environment (which in this case is density defined).

This work puts the investigation of  $L^*$  galaxies into an observational cosmological context by using data from the Galaxy and Mass Assembly project (GAMA: Driver et al. 2011). GAMA is a multi-wavelength photometric and spectroscopic survey, and is designed to answer questions about how matter has assembled on a huge variety of scales: galaxies, groups, clusters and filaments (i.e. all of the environmental scales outlined above). The first phase of the redshift survey was conducted on the AAT (known as GAMA-I) and these data are used in this work (Driver et al. 2011).

GAMA aims to describe the physics occurring in dark matter haloes across the largest mass range ever probed within a single survey, spanning at least 3 orders of magnitude in mass:  $\sim 10^{12} < M_{\text{halo}} < 10^{15} h^{-1} M_{\odot}$ . The GAMA galaxy group catalogue (G<sup>3</sup>Cv1) was recently created using the first 3 years of GAMA-I redshift data and is described in detail in Robotham et al. (2011, hereafter R11). This catalogue was constructed using an adaptive Friends-of-Friends (FoF) grouping algorithm that was extensively tested on mock simulations. The G<sup>3</sup>Cv1 contains LG mass haloes out to  $z \sim 0.1$ , allowing us to better understand how typical MW type galaxies are within a cosmologically significant volume of the Universe.

This work has been divided into multiple papers. The first paper investigated the specific occurrence rates of Magellanic Cloud mass galaxies around MW mass ( $L^*$ ) galaxies – this is a novel question in its own right since much cosmology depends upon the typicality of the Magellanic Cloud satellites of the MW (Robotham et al. 2012). This second paper investigates how the defining characteristics of  $L^*$  galaxies [type, morphology, star formation rates (SFRs), halo mass and stellar mass among others] vary within the total galaxy population, grouped galaxies and close pair systems. This will answer questions on what physics are driving the characteristics of  $L^*$  galaxies like our own MW. The final planned paper will measure the luminosity function of MW mass haloes, revealing how typical the dwarf galaxy population of the MW is.

There has been much written in the literature about the different physical processes that have a measurable effect on the properties of a galaxy. Recent key observational discussions regarding the influence of environment on the properties of galaxies include Baldry et al. (2006), Weinmann et al. (2006), Peng et al. (2010), Thomas et al. (2010) and Kauffmann et al. (2013). The complicating factor has always been in separating out the clear primary role that the local environment has in producing a particular distribution of galaxies (e.g. luminosity or stellar mass: Robotham et al. 2006; Weinmann et al. 2006) and second-order processes that affect properties of these galaxies (e.g. colour and morphology: Driver et al. 2006; Kelvin et al. 2012).

The major roles that a group (or halo) has in assembling stellar mass are: feeding gas into galaxies to enhance star formation (Keres et al. 2005; Dekel & Birnboim 2006); heating and stripping gas out of galaxies (e.g. Gunn & Gott 1972; Abadi, Moore & Bower 1999; Balogh, Navarro & Morris 2000; Dekel & Birnboim 2006) or stripping of stars directly (Bekki, Couch & Drinkwater 2001) and providing a location for rapid galaxy–galaxy interactions (Barnes & Hernquist 1992a), be they harassment, minor or major merger processes (e.g. Moore et al. 1996; Cole et al. 2000, respectively). All of these processes provide mechanisms for stellar mass build-up in galaxies to be either accelerated or decelerated, and will have noisy average net effects as a function of stellar mass, halo mass and epoch. Except for the tidal dwarf variety of galaxies (Barnes & Hernquist 1992b) all galaxies are assumed to be located within a dark matter halo of some description, though to date no rotation curve has been measured for a ‘tidal dwarf’ that does not require some ‘dark’ component (Bournaud et al. 2007). We can therefore say that galaxy stellar mass is entirely coupled to environment, which is not to say the coupling is trivial. The fact it is not trivial has invited a large variety of feedback models in the literature (e.g. Benson et al. 2003; Bower et al. 2006; Croton et al. 2006). Beyond this, it is less clear-cut what additional role the environment has on the observed properties of galaxies.

The second-order properties of galaxies (those things that are not stellar mass) can be largely divided into two categories: luminous energy characteristics (any aspect of a galaxy defined by light output as a function of wavelength: SFR and galaxy colour, etc.) and orbital energy characteristics (aspects concerning the position and velocity of stars: morphology and colour gradients, etc). Processes that can affect the luminous energy characteristics are mostly those that directly enhance or suppress star formation (which is then necessarily coupled to the stellar mass build-up discussed above) and more contentiously any process that affects the initial mass function (IMF) that stars are forming with (see Gunawardhana et al. 2011, for recent work on variable IMFs).

Orbital properties within galaxies can evolve naturally through ‘secular’ processes that allow for orbital energy redistribution, which we observe as brief and varying manifestations such as pattern bars and spiral arms (e.g. Sellwood & Wilkinson 1993; Bertin et al. 1989, respectively), or permanent transitions such as the formation of thick discs and pseudo-bulges (e.g. Haywood 2008; Kormendy & Kennicutt 2004, respectively). Orbital properties are also able to be changed more dramatically and violently during encounters with other galaxies (Barnes & Hernquist 1992a). Minor merger events can produce subtle changes to stellar orbits, providing another mechanism for the creation of a thick disc (Read et al. 2008). Major mergers can be incoherent enough to transform a galaxy with regular rotational disc-like stellar structure (the canonical late-type galaxy) into a system with random orbits (the canonical early-type galaxy).

Self-evidently ‘secular’ processes are not a direct consequence of environment since they describe processes that should occur regardless of outside influence/trigging; therefore, they might be expected to be entirely predictable from the instantaneous properties of a galaxy. However, environment can play a role in disrupting secular processes, e.g. a major merger would destroy any bar formation process. Clearly there is scope for environment to play at least a secondary role in almost all aspects of galaxy evolution, if only by virtue of disruption, delay and prevention mechanisms. The purpose of this paper is to simultaneously separate out the stellar mass, galaxy–galaxy interaction, and halo mass terms in order to quantify the codependencies of secularly and environmentally driven stellar mass evolution.

Secondary properties of galaxies clearly display differing sensitivity to modifying external drivers. E.g.  $H\alpha$  equivalent width (EW) is a relatively instantaneous response to an increase in star formation, accompanied by a change in integrated bandpass colour to bluer colours. In the longer term the in situ gas and locked-up stellar metallicity will increase, which will create a redder spectral energy distribution (SED). If this burst of star formation occurred due to a minor merger event then we might also expect orbits to become disrupted, leading to a change in the velocity dispersion and light profile of the galaxy. If the change is dramatic enough (e.g. a major merger) it might also invoke a change in the ‘type’ classification of the galaxy. Thus, we have galaxy properties that are highly sensitive to current star formation invocations ( $H\alpha$  EW) and ones that betray more about the integrated history of a galaxy (some component of colour and all aspects of morphology). Depending on the physical processes that caused star formation we might also observe longer term changes to the morphology of the galaxy (in the case of minor or major merger events). In this work we aim to separate out these different secondary properties into those most dependent on different likely primary drivers of galaxy properties: stellar mass, halo mass and galaxy–galaxy interactions.

In Section 2 we discuss the data used in this work in detail. In Section 3, we discuss the sample selections applied to the data. In Section 4, we describe the visual classification process that was used to categorise the morphological type, emission class and presence of a disc. In Section 5, we discuss the different varieties of galaxy categorisation used for constructing later plots. In Section 6, we analyse how properties of MW mass galaxies are affected both by the group-scale and by the pair-scale environment, where the codependence of galaxy properties is discussed in detail. In Section 7, we pull together the main results of Section 6 into a coherent narrative and put the findings into context. In Section 8, we briefly summarize the major findings of the work.

In this work, we use  $\Omega_M = 0.25$ ,  $\Omega_\Lambda = 0.75$  and  $H_0 = 70 \text{ km s}^{-1}$ . This is consistent with the first paper in this series that dealt specifically with the commonality of Magellanic Cloud galaxies around  $L^*$  systems (Robotham et al. 2012).

## 2 DATA

### 2.1 GAMA

GAMA is a major new multi-wavelength photometric and spectroscopic galaxy survey (Driver et al. 2011). The final redshift survey will contain  $\sim 300\,000$  redshifts to  $r_{AB} = 19.8$  over  $\sim 280 \text{ deg}^2$ , with a survey design aimed at providing an exceptionally uniform spatial completeness (Baldry et al. 2010; Robotham et al. 2010; Driver et al. 2011).

Extensive details of the GAMA survey characteristics are given in Driver et al. (2011), with the survey input catalogue described in Baldry et al. (2010) and the spectroscopic tiling algorithm in Robotham et al. (2010). The first 3 years of data obtained has been frozen for internal team use and is referred to as GAMA-I.

Briefly, the GAMA-I survey covers three regions each  $12 \times 4$  deg<sup>2</sup> centred at 09<sup>h</sup>, 12<sup>h</sup> and 14<sup>h</sup>30<sup>m</sup> (respectively, G09, G12 and G15 from here). The survey depths and areas relevant for this study are:  $\sim 96$  deg<sup>2</sup> to  $r_{\text{AB}} = 19.4$  mag (G09 and G15) and  $\sim 47$  deg<sup>2</sup> to  $r_{\text{AB}} = 19.8$  (G12).<sup>1</sup> All regions are more than 98 per cent complete (see Driver et al. 2011, for precise completeness details), with special emphasis on a high close pair completeness, which is greater than 95 per cent for all galaxies with up to five neighbours within 40 arcsec of them (see fig. 19 of Driver et al. 2011). Stellar masses used for this work are taken from Taylor et al. (2011).

## 2.2 Spectral analysis

The spectra collected during the GAMA observations have been flux calibrated, Balmer decrement corrected, aperture corrected, and the emission-line EWs measured. The spectral pipeline process is discussed in detail in Hopkins et al. (2013). The line measurements have also been converted into SFRs, a process that is described in Gunawardhana et al. (in preparation). These SFRs are used directly in this work.

These emission line measurements are also used for automatic spectral classifications by comparing the  $[\text{O III}]/\text{H}\beta$  as a function of  $[\text{N II}]/\text{H}\alpha$  in the popular manner outlined in Baldwin, Phillips & Terlevich (1981), widely known as BPT classification. This process, and the classifications selections used, are discussed in detail in Hopkins et al. (2013) and Gunawardhana et al. (in preparation). These automatic classifications are supplemented by extensive visual checks to minimize classification errors (discussed below).

For this work we keep the classifications as simple as possible:  $\text{H}\alpha$  emission that is dominated by star formation irradiation,  $\text{H}\alpha$  emission that is dominated by active galactic nucleus (AGN) irradiation and no  $\text{H}\alpha$  emission. This binary definitions of star formation means we are highly insensitive to the details of flux calibration, Balmer decrement correction and aperture correction for the majority of our work. The fidelity of the spectral classifications is investigated below.

## 2.3 G<sup>3</sup>Cv1

One of the principal science goals of GAMA is to make a statistically significant analysis of low-mass groups ( $M \leq 10^{13} M_{\odot}$ ), helping to constrain the low-mass regime of the dark matter halo mass function and galaxy formation efficiency in LG-like haloes. This low-mass regime contains groups that have halo masses similar to the LG, allowing us to investigate how common the LG environment is, and putting LG science into a cosmological context.

The GAMA-I spectroscopic data were used to construct the first GAMA galaxy group catalogue (G<sup>3</sup>Cv1: R11). This group catalogue was constructed using an adaptive Friends-of-Friends (FoF) algorithm that was optimised to recover groups in the magnitude limited GAMA-I spectroscopic data.

In total G<sup>3</sup>Cv1 contains 14 388 groups (with multiplicity  $\geq 2$ ), including 44 186 galaxies out of a possible 110 192 galaxies, im-

plying  $\sim 40$  per cent of all galaxies are assigned to a group. Since GAMA-I has an uneven  $r$ -band selection function between GAMA regions we use an  $r < 19.4$  mag subset of the full G<sup>3</sup>Cv1 for this work. This catalogue has 37 576 galaxies in 12 200 groups.

The G<sup>3</sup>Cv1 contains a wealth of ancillary data regarding the properties of the groups, including dynamically estimated halo masses. These halo masses are used extensively in this work. For details regarding the reliability and biases, the reader should refer to R11 and Alpaslan et al. (2012).

## 2.4 Pair catalogue

This paper is investigating the effect that various kinds of environment have on the characteristics of  $L^*$  galaxies. Close galaxy interactions have a significant role on the evolution of galaxies, so these systems are selected for close analysis in this work. Part of the process of creating the G<sup>3</sup>Cv1 catalogue involves the construction of all galaxy pairs (see R11 for details). This catalogue contains galaxies with potentially quite large radial (velocity) and tangential (spatial distance) separations between galaxies. Using the full pair catalogue would include galaxy pairs with very large dynamical times, and such pairs will have weak evolutionary effects on the component galaxies. Instead we select a narrow window of interaction phase space in order to preferentially extract galaxies that will be most affected by close galaxy–galaxy interactions. The ‘close pair’ sample selected here is based on that presented in Robotham et al. (2012), where we aimed to recover galaxy pairs that are similar to the MW Magellanic Clouds system. This ensures we can investigate a variety of interaction that our own MW  $L^*$  galaxy has experienced, helping to place it in the broader family of  $L^*$  galaxies.

The Small Magellanic Cloud (SMC) is  $\sim 60$  kpc from the MW and travelling radially away at  $\sim 17$  and  $301$  km s<sup>−1</sup> tangentially with respect to the MW: a net velocity of  $302$  km s<sup>−1</sup> (Nichols et al. 2011). The Large Magellanic Cloud (LMC) is  $\sim 50$  kpc from the MW and travelling radially away at  $\sim 89$  radially and  $367$  km s<sup>−1</sup> tangentially with respect to the MW: a net velocity of  $378$  km s<sup>−1</sup> (Nichols et al. 2011). To conservatively recover all systems where the galaxies are in such spatial–velocity proximity, we create a catalogue of pairs for this work where the projected separation is  $r_{\text{sep-proj}} < 70$  kpc and the radial velocity separation is  $v_{\text{sep-rad}} < 400$  km s<sup>−1</sup>. See Robotham et al. (2012) for detailed discussion on galaxy pair selections.

Applying these selection limits to recover all systems that have similar pairwise dynamical properties to the MW and the Magellanic Clouds creates a catalogue containing 5964 galaxy pairs and 10 416 unique galaxies. These numbers are for the whole GAMA-I data set, so include stellar masses that might be highly dissimilar to the real MW Magellanic system.

Some galaxies will be members of more than one pair (the MW is paired with two galaxies – the SMC and LMC). We create groupings that contain all galaxy–galaxy associations and count this as a single ‘pair’ system, e.g. the MW is in a pair system of three, as are the SMC and LMC. In the same way that simple galaxy pairs are only counted once, higher order pair systems are also only counted once. Table 1 shows the frequency of different pair systems of galaxies. While we recover 5964 galaxy pairs, once common links are explored we are left with 5446 systems.

## 3 SAMPLE SELECTION

A large component of the analysis required for this paper required comprehensive visual classification (described in detail below). The task of visually classifying the data in this manner was extremely

<sup>1</sup> See Baldry et al. (2010) for additional GAMA-I selections.

**Table 1.** Number of pair systems with different numbers of galaxies.

Galaxies in paired systems	Frequency	Total galaxies
2	4550	9100
3	370	1110
4	45	180
5	4	20
6	1	6

labour intensive – each galaxy could take a minute to analyse and GAMA-I contains 129 765 galaxies with redshifts above 0.01. Doing this for 8 h a day would take 270 d. Instead we selected a subset of galaxies concentrated on  $L^*$  systems (the focus of this paper), and systems that share common properties with the MW and its local environment (allowing us to identifying the evolutionary drivers for the MW).

Since this work is predominantly interested in the properties of MW mass ( $L^*$ ) galaxies we had to comfortably include all GAMA galaxies with stellar mass similar to the MW ( $5.0 \times 10^{10} M_{\odot}$ ; Flynn et al. 2006). A conservative lower stellar mass limit of  $\mathcal{M}_S = 10^{10} M_{\odot}$  was chosen. Due to the rapid decrease in the number of objects as we move into the higher stellar mass regime there was no need to implement a higher stellar mass cut-off. In total this stellar mass selected sample contains 3550 galaxies.

Since we are also investigating the environment that  $L^*$  galaxies reside in, we add all galaxies that are associated with the  $L^*$  selection to the sample. There are two varieties of ‘environment’ investigated in this work: galaxy pairs and galaxy groups. These are just two of an infinite number of possible definitions of environment, but they probe the smallest physical scale environments possible, and as such should possess the cleanest possible relationship between environment and galaxy properties due to the recent dynamical events. To link this work closely with our own MW environment (important since so much effort is being invested in surveys of the MW and local galaxies), we are careful to ensure that our definitions of ‘pair’ fully capture interactions like the MW is undergoing with the LMC (its dominant pairwise interaction). We also select all groups that are similar mass to the MW halo, allowing us to investigate how common the MW system is, given its halo mass as a prior.

First we add our pair selection criteria. We include all pairs within a redshift limit of  $z = 0.089$  that have at least one member selected in the mass sample described above. This allows us to observe LMC equivalent galaxies throughout our sample (see Robotham et al. 2012, for details of the selection limits). Together with the lower redshift limit of  $z = 0.01$  this gives a sample of 3,503 galaxies. In this work we use stellar mass rather than  $r$ -band magnitude to describe the galaxies, and the redshift limit  $z = 0.089$  is specifically the implied  $r$ -band magnitude limit to observe LMC type galaxies. However, since the LMC is blue, late type and star forming, this will act as a very conservative LMC stellar mass limit. Indeed the  $\mathcal{M}_S > \mathcal{M}_{S, \text{LMC}}$  number counts drop off at  $z \sim 0.1$ . For consistency with Robotham et al. (2012) we use the more conservative limit of  $z = 0.089$  to select down to LMC mass galaxies.

In order to determine what role different environments may have on the evolution of  $L^*$  systems, we then select all galaxies that are in any  $G^3\text{Cv}1$  group or galaxy–galaxy pair containing any of the 3,503 galaxies in the above subset. This gives us an expanded sample of 5,054 galaxies. We add to this all galaxies in all groups that have a  $G^3\text{Cv}1$  halo mass between  $7 \times 10^{11} < M_{\text{FoF}} < 10^{13} M_{\odot}$  where  $M_{\text{FoF}}$  is defined using the functional form of the halo mass used

in R11. This halo mass range comfortably covers both sides of the MW halo mass used in Robotham et al. (2012,  $2.5 \times 10^{12} M_{\odot}$ ) and the total LG mass (MW+M31 haloes =  $\sim 5 \times 10^{12} M_{\odot}$ ; Li & White 2008). Including this halo mass range guarantees that all systems with properties similar to the MW by virtue of stellar mass, and halo mass or close pair interactions, are certain to be included in our sample. This gives us a final sample of 6150 galaxies to visually classify in detail. As long as the stellar mass and halo mass ranges are kept within the limits described above, this sample is volume limited.

The result is our sample has two major components:

- (i) All GAMA galaxies with  $\mathcal{M}_S \geq 10^{10} M_{\odot}$  (mass sample, with 3550 galaxies).
- (ii) Within  $0.01 < z < 0.089$ , all pairs and groups are included that have at least one member in the above mass selection. These two varieties of environment are therefore complete down to LMC mass galaxies. In addition all  $G^3\text{Cv}1$  groups with a halo mass in the range  $7 \times 10^{11} < M_{\text{FoF}} < 10^{13} M_{\odot}$  and  $0.01 < z < 0.089$  (volume sample, with 6150 galaxies).

Plots that only require a mass selection are defined by the mass sample. Any plot that makes use of pair or group information is defined by the second volume limited selection. By using this combination of stellar mass, group and pair selections we are able to investigate all of the key characteristics that define the MW and the LG environment.

#### 4 VISUAL INSPECTION OF THE DATA

In this work, we are primarily interested in how the presence of a group environment, or another nearby galaxy, can affect the characteristics of a given stellar mass galaxy, particularly in the regime of MW mass galaxies (i.e.  $L^*$  galaxies).

To undertake such analysis it is necessary to know various characteristics about  $L^*$  mass galaxies (i.e. the knee of the stellar mass function,  $\mathcal{M}^*$ ). Some of these characteristics can be determined directly from the data: colour, mass. Others are more difficult to extract from the data due to the subjective nature of the variable or the imperfect quality of the data.

Galaxy type, by which we mean morphologically early type or late type, is difficult to determine from the catalogue data directly. When analysing a fiducial sample of galaxies taken from Kelvin et al. (2012) the relationship between Sérsic index and morphological type is highly imperfect: it is not true to say that high Sérsic index ( $n > 3$ ) galaxies are early type and low Sérsic index ( $n < 3$ ) galaxies are late type. Inclination in particular can distort catalogue values, and the reliability of colour as a galaxy ‘type’ discriminant is heavily compromised by the presence of dust.

Star formation rate [or specific SFR (sSFR)] appears to be a characteristic that is well suited to catalogued data. However, analysis of a fiducial sample revealed that the current emission line based SFRs (Hopkins et al. 2013) fail to detect SFRs for  $\sim 20$  per cent of galaxies where the emission line is very weak and the spectrum has low signal-to-noise ratio (S/N). The lack of a measurement is not an error, rather a pragmatic decision to improve the reliability of quoted SFRs by ignoring marginal lines. A particular problem is that these undetected weakly star-forming galaxies are typically fainter lower mass systems, creating a stellar mass bias in our analysis of star formation. Also, at present GAMA incorporates spectra from non-AAOmega sources (e.g. SDSS, 2dFGRS) and emission-line analysis has not yet been run on the entire sample. This has a larger effect at lower redshifts, which is the regime of interest in

this work, since a larger fraction of spectra are non-AAOmega for brighter magnitudes. Ultimately, independently identifying spectral emission features visually was deemed to be an appropriate route forward for the science investigated in this paper, where we are mostly interested in the binary divide of star formation: on or off. This route of investigation is informed by recent studies (e.g. Wijesinghe et al. 2012) showing the star-forming fraction, not the rate of star formation, is the property sensitive to the ‘environment’. This requires us to push the usable S/N limits of the catalogues, which is discussed in more detail below.

The presence, or otherwise, of a disc is also a key characteristic that is difficult to extract from catalogued data. Using the *data-model* residual images produced by Kelvin et al. (2012) it was clear that a large number of early-type galaxies, even with high Sérsic indices, possess weak but distinct discs. While late-type galaxies, by their definition, always possess a disc-like structure, there exists a large fraction of early-type galaxies where this might be a key distinguishing variable (we include S0 and lenticulars in this class).

To determine classes for these three categories a large subset of the GAMA-I survey was studied in detail by ASGR. Using the single object viewer (SOV) tool developed by JL it was possible to view the following information simultaneously for each galaxy:

- (i) colour image;
- (ii) basic catalogue information;
- (iii) preferred spectrum;
- (iv) 2D modelling information, including difference image;
- (v) 1D surface brightness profile.

Fig. 1 is a screen grab of the SOV tool for GAMA galaxy 6926. Using the information available in the SOV, ASGR assessed the following three galaxy characteristics:

- (i) whether the galaxy is late type (visibly disc dominated) or early type (visibly bulge dominated);
- (ii) whether the galaxy has any evidence for a disc (binary true or false);
- (iii) whether the galaxy exhibits star-forming H $\alpha$  emission, no H $\alpha$  emission or is an obvious AGN (broad emission features).

The three classifications used to quantify emission class are shown in Fig. 2. As well as assessing galaxies for these different classes, data quality was simultaneously checked. Galaxies were assessed as to whether they were sub-structure of larger galaxies or misclassified stars that have a redshift virtue of background galaxy flux. 16 galaxies were re-classified as substructure and six galaxies were re-classified as stars, i.e. 99.64 per cent of galaxies in our original sample were correctly identified as being galaxies (see Baldry et al. 2010, for details of the original GAMA star galaxy separation).

Finer resolution grading of properties was not possible over the redshift range probed, i.e. it was not possible to reliably identify the Hubble class or similar lower order classifications. The properties assessed were subjectively considered to be the most information that could be robustly extracted given the quality of data. To guide later discussions, lenticular galaxies and early-type spiral galaxies would both be classified as early-type galaxies possessing discs.

## 5 GALAXY CATEGORIZATION

For multi-dependent data mosaic plots are a very powerful and information rich means of visualization. They differ from principal component analysis (PCA) in that they readily handle discrete data classes, which PCA does not. For readers unfamiliar with how mosaic plots are constructed, and best interpreted, Appendix A de-

scribes them in detail. It is strongly recommended that novices to the mosaic plotting technique start there. To ease the comprehension of the large number of plots presented, the major observations drawn from each figure are included in the relevant captions. This should be particularly helpful in explaining the mosaic plots to those unfamiliar with them.

In order to present information in the form of mosaic plots, all information must be discretized. Some categories are naturally discrete: morphological type and presence of disc. Others have to be coerced into discrete categories: SFR (as outlined above). Here we describe all of the galaxy categorizations that are used to construct later mosaic plots, along with the abbreviation in parentheses used for labelling.

### 5.1 Naturally discrete galaxy categories

Below are listed all of the naturally discrete classes of galaxy properties investigated. The determination of these divisions require less subjectivity than for continuous data. However, not all categories are fully objective – opinion is still required to determine whether a galaxy is early type or late type.

#### 5.1.1 Galaxy type

Early Type (e)/Late Type (l)

As part of the visual classification process outlined in Section 4, galaxies are categorised by simple Hubble type. The divide used was whether the galaxy appeared to be bulge dominated ( $R_e$  inside the bulge regime, which is determined by a strong visual break in the 1D light profile) or disc dominated (either no bulge, or  $R_e$  outside the bulge regime). This estimation is informed by both multi-band and  $r$ -band SDSS images, and the 1D profile of 2D single Sérsic fits generated by LSK (see Kelvin et al. 2012, for details on the single Sérsic fitting process). Further resolution beyond early-type and late-type is not attempted due to the quality of the data. Lenticulars (S0) galaxies would be classified as early type in this scheme, as would early-type spiral galaxies.

#### 5.1.2 Galaxy disc

False (F)/True (T)

As part of the visual classification process outlined in Section 4, galaxies are assessed to either possess some form of disc, or not. This subjective observation is based on the  $r$ -band Sérsic profile subtracted residual images created by LSK (see Kelvin et al. 2012, for details on the single Sérsic fitting process). Further detail beyond this binary status is not attempted due to the quality of the data. Lenticulars (S0) galaxies would be classified as possessing a disc in this scheme.

#### 5.1.3 Galaxy in group

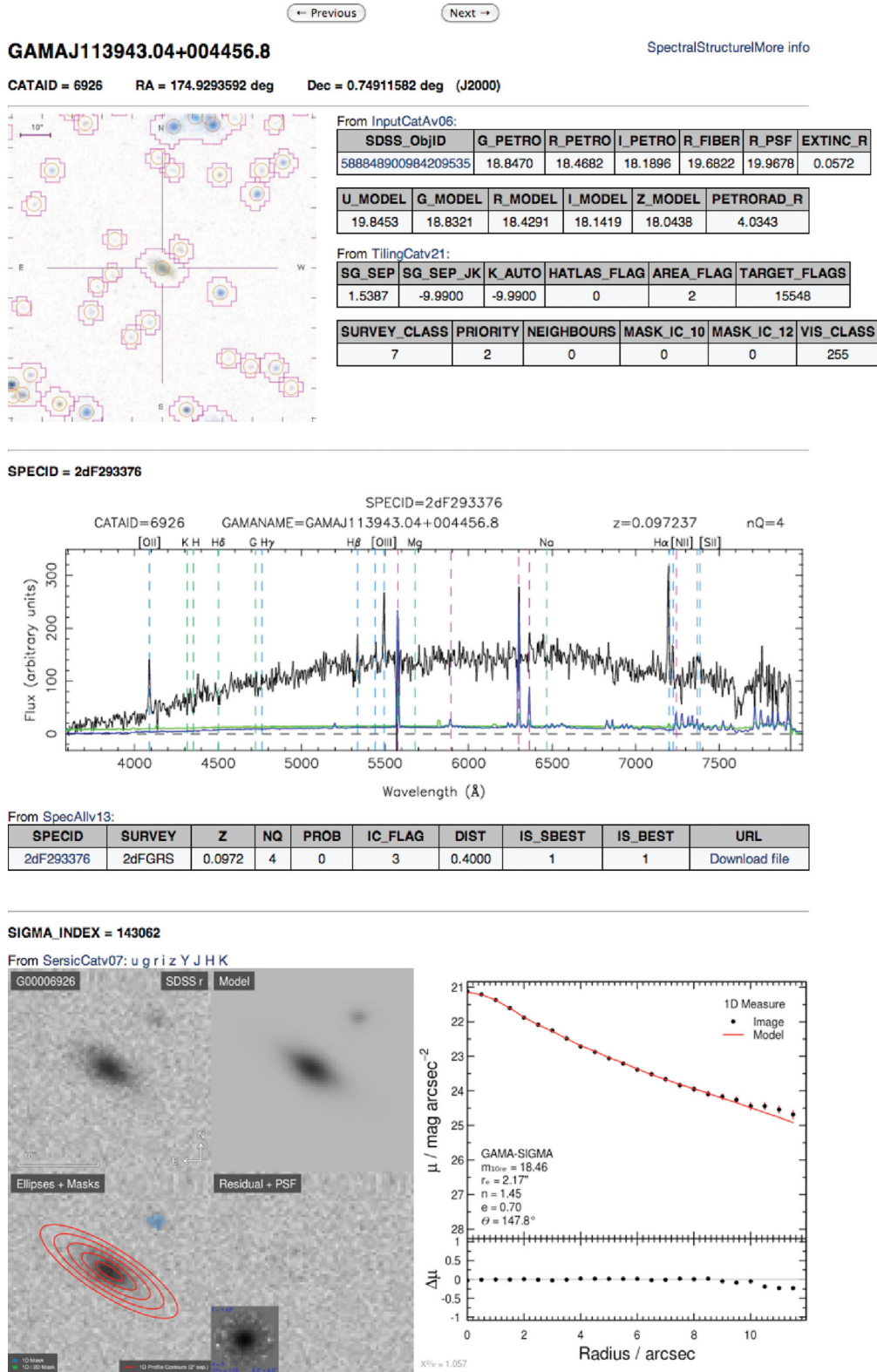
False (F)/True (T)

Logic class defining whether a galaxy is found within a G<sup>3</sup>Cv1 group that makes the selection criteria outlined in Section 3.

#### 5.1.4 Galaxy is central or satellite

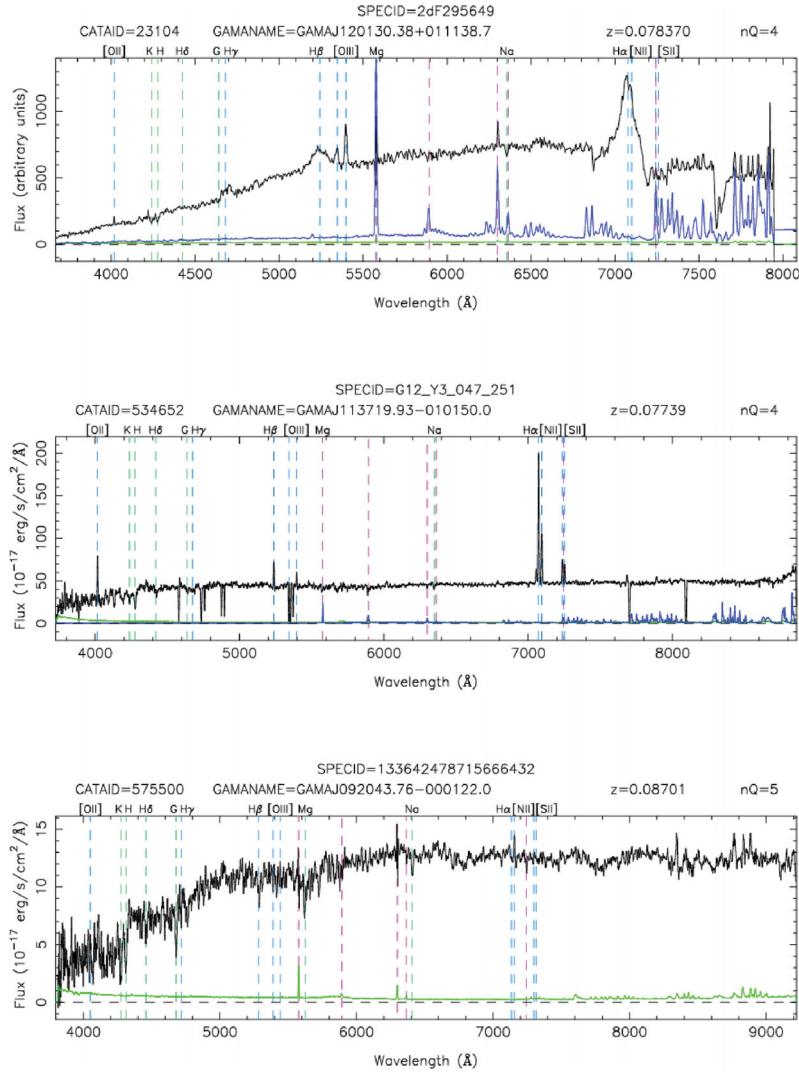
Central Galaxy (C)/Satellite Galaxy (S)

Class defining whether a galaxy is the largest (in terms of stellar mass) in its group, otherwise referred to in this work as the biggest



**Figure 1.** Screen grab of a typical display using the SOV tool available for the GAMA data base. In the top left-hand side, there is an image taken directly from the SDSS image server. In the top right-hand side, there are some key catalogue values for the current galaxy. In the middle is the preferred GAMA spectrum with positions of emission and absorption lines over-plotted. Directly below the spectrum is a list of other available spectra in the GAMA data base (useful when there are data quality issues in key regions of the preferred spectrum). In this example, the only available spectrum is from 2dFGRS. Bottom-left is the image output of Kelvin et al. (2012). Within this panel, the top left-hand side is the *u*- to *K*-band image (*r*-band is the default, but others can be selected by clicking on the links above), top right-hand side is the final Galfit model, bottom left-hand side shows the ellipses used to construct the 1D surface brightness profile and bottom right-hand side is the *data-model* residual image with the local PSF overplotted. The bottom right-hand side of the whole figure is the 1D surface brightness profile estimated using IRAF ELLIPSE, with key values regarding the fit printed inside the plot window.





**Figure 2.** Examples of the three emission classes used for visual classification. The top panel shows broad line features due to high velocity emission from an AGN (class A). The middle panel shows high-EW  $H\alpha$  emission relative to all other emission lines, so a star-forming spectrum (class S). The bottom panel shows no signs of any  $H\alpha$  emission on top of the continuum (X).

group galaxy (BGG). If it is then it is defined as the ‘central’ galaxy (C), otherwise it is a ‘satellite’ (S). All ungrouped galaxies are defined as central galaxies in their group of multiplicity 1 (but these 1 member groups do not count as part of the  $G^3Cv1$ ).

### 5.1.5 Galaxy in a pair system

False (F)/True (T)

Logic class defining whether a galaxy found within a  $G^3Cv1$  pair system that makes the selection criteria outlined in Section 3.

### 5.1.6 Galaxy is the largest in a pair system

False (F)/True (T)

Logic class defining whether a galaxy is the largest (in terms of stellar mass) in its pair system.

## 5.2 Naturally continuous galaxy categories

Below are listed all of the naturally continuous classes of galaxy properties investigated. The determination of these divisions contain

some subjectivity on our part. Making continuous data discrete obviously reduces the information content available to us, but in practice it is often the only way to visualize behaviour of noisy relationships: it is common practice to plot the running median in scatter plots, which is a discrete separation of the data. Even binning data when plotting a luminosity function is a discretizing process. Obviously any further numerical analysis should make use of the underlying continuous data.

### 5.2.1 Emission strength

AGN (A)/Star Forming (S)/No Emission (X)

As part of the visual classification process outlined in Section 4 galaxies are visually assessed into these three classes of emission. AGN classes (A in plots) are those with clear  $H\alpha$  emission, but where the spectral signature is dominated by AGN irradiation. Star-forming classes (S in plots) are those with clear  $H\alpha$  emission, but where the spectral signature is dominated by star formation irradiation. No Emission classes (X in plots) are those without any  $H\alpha$

emission, though other lines may still be present (typically weak  $[\text{N II}]$ ).

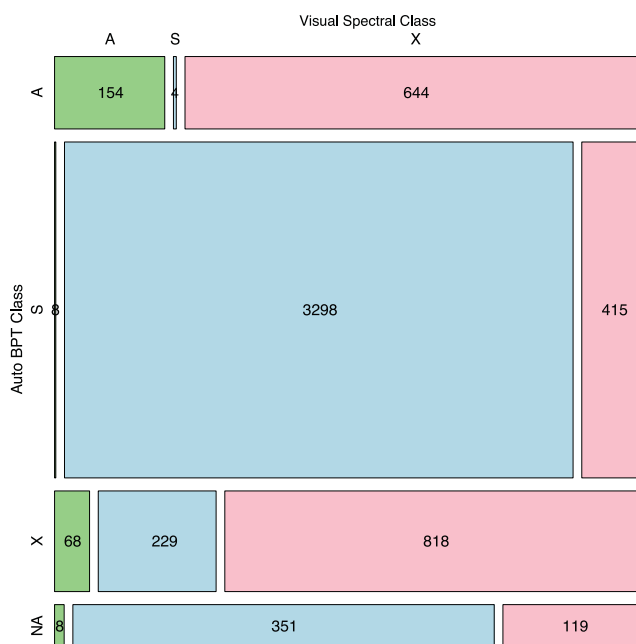
Fig. 2 gives an example of each. Galaxies were both spectrally classified through a classic BPT separation (Baldwin et al. 1981) and by eye. The GAMA application of the BPT separation is described in detail in Gunawardhana et al. (in preparation) and Hopkins et al. (2013), and includes all GAMA and SDSS origin spectra (which covers 92.2 per cent of the sample investigated here). In the automatic classification scheme, a galaxy is classed as ‘X’ when it cannot be placed on a BPT, i.e. it is missing at least one of the  $\text{H}\alpha$ ,  $\text{H}\beta$ ,  $[\text{O III}]$  or  $[\text{N II}]$  lines. This differs to the definition desired in this work, that ‘X’ refers only to systems with no  $\text{H}\alpha$  emission. By using this definition the separation between classes is much clearer.

Separately to this purely automatic process, all spectra were classified by eye. It is easy to identify AGN spectra when there are broad emission line features, but these are a minority of cases. In other cases the decision was made by dividing the height of the  $[\text{N II}]$  and  $\text{H}\alpha$  features, and classifying the galaxy as AGN when this ratio was clearly above 0.5. This uses part of the information available in a formal BPT analysis, and will be approximate at best. As such, a clear weakness of visually classifying spectra is separating marginal BPT classifications. All disagreements were followed up at least once and a discernment made between cases where the automatic process was distinguishing between subtle emission ratios that cannot be accurately determined by eye, and those where an fitting error has occurred, creating a poor automated fit.

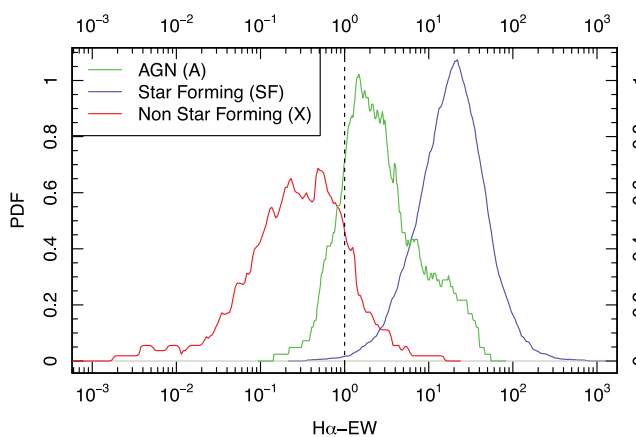
In cases where the BPT was able to select marginal AGN systems and  $\text{H}\alpha$ ,  $\text{H}\beta$ ,  $[\text{N II}]$  and  $[\text{O III}]$  were all in emission, the BPT was preferred. However, the majority of BPT selected AGN classifications were tenuous low powered LINER-like systems with no significant  $\text{H}\alpha$ ,  $\text{H}\beta$  or  $[\text{O III}]$  flux. In these situations only a weak  $[\text{N II}]$  line was present (as in the bottom panel of Fig. 2), and the classification was changed to ‘X’, since there is no strong  $\text{H}\alpha$  flux. This is the main source of disagreement revealed during visual classification of spectra. It is important to change these classifications since a significant number of AGNs are, by the definitions we wish to use, erroneous.

Fig. 3 summarizes the BPT versus visual classification. Overall 75.4 per cent of automated BPT classifications agree with the visual classifications. For AGNs the figure drops to 19 per cent for the reasons described above (only exceptionally obvious AGN can be reliably picked out by eye). For star-forming systems the agreement is high, at 88 per cent. Finally, the non-emission classes agree at the 73 per cent level. In addition, a non-negligible (7.8 per cent) fraction of sources do not have SDSS or GAMA spectra; therefore, they do not have a BPT classification in the GAMA data base (see the bottom panels of Fig. 3). These systems had to be classed entirely by eye, but broadly reflect the overall classification distributions seen for the full sample. The main survey source for these galaxies without an automated BPT classification is 2dFGRS.

Fig. 4 demonstrates the  $\text{H}\alpha$ -EW distributions for the final classifications used in this work. Star-forming galaxies have a typical  $\text{H}\alpha$  EW at least a factor 10 and on average a factor 100 larger than the visually identified non-star-forming galaxies. The majority of non-star-forming galaxies do not have any measured  $\text{H}\alpha$  flux (83.8 per cent). This plot shows the probability density functions (PDFs) for the small fraction that have some flux detected during the automated line measuring conducted for GAMA (see Hopkins et al. 2013, for details). An approximate version of our classifications can be made using  $\text{H}\alpha$ -EW alone: 97.6 per cent of galaxies that have an  $\text{H}\alpha$ -EW above 1 are either AGN or star forming, and 93.3 per cent of galaxies with  $\text{H}\alpha$  EW below 1 (including non-measurable) are non-star forming. Using  $\text{H}\alpha$  EW alone does not allow for a clean separation



**Figure 3.** Mosaic plot comparing the automated BPT classifications (horizontal bands) versus the visual classifications (vertical coloured bands). For both classifications A = AGN, S = star forming and X = non-star forming. For the visual classifications, green = AGN, blue = star forming and light red = non-star forming. The most disagreement comes for AGN classes (A) where very low powered LINER-like systems with no visible  $\text{H}\alpha$  are being classed as AGN. Since we are most interested in  $\text{H}\alpha$  emission, and only wish to exclude cases where a large component of the  $\text{H}\alpha$  flux is due to AGN activity, these objects are mostly moved to the no emission class (X). Spectra taken from surveys other than GAMA and SDSS do not have an auto classification (shown as NA). For a detailed description of how to interpret mosaic plots, the reader should refer to the appendix. Observations: strong agreement for ‘S’ class emissions; large fraction of auto AGN classes have no  $\text{H}\alpha$  emission.



**Figure 4.** Comparison of the  $\text{H}\alpha$ -EW PDFs for the AGN (A), star-forming (S) and non-star-forming (X) classifications used in this work, where the  $\text{H}\alpha$  emission has been measured during the GAMA line measuring process (see Hopkins et al. 2013, for details). As expected, where  $\text{H}\alpha$  has been measured the EW is significantly higher for star forming compared to non-star-forming galaxies, with AGN classed galaxies sitting broadly between the two. To aid comparison with other work, we suggest dividing the  $\text{H}\alpha$ -EW at 1 (vertical dashed black line) and classing galaxies above this threshold as AGN/star forming and classing galaxies below this threshold as non-star forming.

of AGN and star-forming galaxies (hence the popularity of BPT diagrams and similar schemes), but these divisions will allow some of the more dramatic results to be compared to other work.

The AGN classifications contain within them complex subclasses that we do not attempt to resolve in this work: i.e. QSO, Seyfert type 1 and 2, and strong LINER galaxies. Since the physical origin of these different classes are quite different and some of the variation is due to observational bias (e.g. distinction between Seyfert classes is largely observational orientation), we will focus this work on the star formation (S) and no emission (X) classes. However, the AGN classes will be listed in all mosaics in order to accurately represent the full population fractions. In our final categorisations, 3.9 per cent of galaxies are AGN with  $H\alpha$  emission, 63.1 per cent have star-forming  $H\alpha$  emission, and 32.5 per cent have no  $H\alpha$  emission. The final 0.5 per cent of objects are not real galaxies, either stars or substructure of extended galaxies.

### 5.2.2 Colour

#### Blue (b)/Green (g)/Red (r)

Colour is an important characteristic of a galaxy. Here we treat colour as a purely relative quantity, and we use  $g - r$  because we have detections in both bands for all galaxies. It is known that the bimodal nature of galaxy colour is enhanced by using a bluer band than  $g$  since this only partly samples blue-wards of the 4000 Å break (Baldry et al. 2004), but since colour does not form the main focus of future analysis, completeness was preferred. Also, since we divide colour into relative quantiles, the lack of  $g - r$  bimodality should not dramatically alter later plots since bluer galaxies do tend to be bluer in both  $X - r$  and  $g - r$ , where  $X$  is any band bluer than  $g$ . The Galactic dust corrected and  $k$ -corrected (using  $k$ -correct and GAMA's  $ugrizYJHK$  photometry, Blanton & Roweis 2007; Hill et al. 2011, respectively)  $g - r$  colour is separated into three equal sized quantiles of data for galaxies within 0.15 dex of  $\mathcal{M}_{S,MW}$ . The limits of the quantiles are used to define galaxies as blue, green or red in a relative sense:

$$g - r < 0.77 = \text{Blue(b)}$$

$$0.77 < g - r < 0.86 = \text{Green(g)}$$

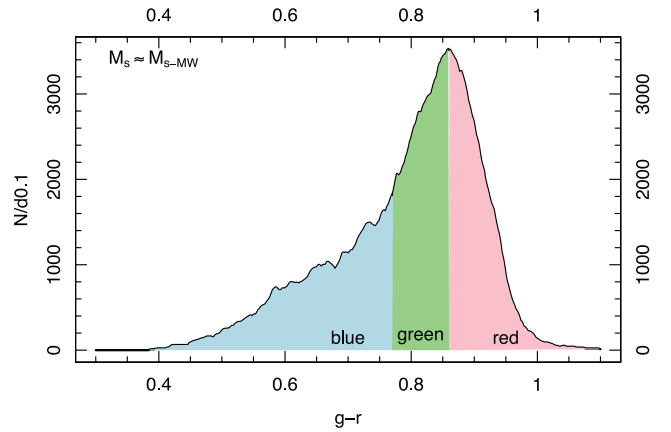
$$g - r > 0.86 = \text{Red(r)}$$

Fig. 5 shows the distribution of  $g - r$  colours for the mass selection used, and highlights the divisions used. There is no clear bimodality for this mass selection and colour combination, although the blue galaxies exist in a distinct wing. The green and red galaxies are largely selecting different sides of the dominant colour distribution peak.

### 5.2.3 Stellar mass

$$X \times \mathcal{M}_{S,MW} \quad (1)$$

We split GAMA galaxy stellar masses (taken from Taylor et al. 2011) into bins separated by factors of two, and each bin therefore contains a mass range of  $\pm 0.15$  dex.  $\mathcal{M}_{S,MW}$  is the reference MW stellar mass. Since we are complete down to  $\mathcal{M}_S = 10^{10} M_\odot$  we define the bin centre of the  $X = 1$  bin containing the MW as  $5.67 \times 10^{10} M_\odot$  (so it includes stellar mass of the MW, which we take to be  $\mathcal{M}_{S,MW} = 5.0 \times 10^{10} M_\odot$ ; Flynn et al. 2006), meaning we have a complete sample with 2.5 bin widths on the low-mass side exactly. The stellar mass range selected is 100 per cent complete for galaxies where  $X \geq \frac{1}{4}$ .



**Figure 5.** The distribution of  $g - r$  colours for galaxies with stellar masses within 0.15 dex of  $\mathcal{M}_{S,MW}$ . The blue/green/red divisions used are shown by the different coloured shading. Selection shown is the mass sample (Section 3). Observations: (i) There is no clear bimodality in the  $g - r$  colour for the MW stellar mass selection used to define the colour quantiles. (ii) The blue population selects the majority of the outlying wing in the  $g - r$  distribution, while the green and red population sample on different sides of the main population peaks. This suggests that the green and red selections are likely to be selecting galaxies driven by similar physical processes.

For reference the MW is in the  $X = 1$  bin ( $\mathcal{M}_{S,MW} = 5.0 \times 10^{10} M_\odot$ ; Flynn et al. 2006); the LMC is in the  $X = \frac{1}{32}$  bin ( $\mathcal{M}_{S,LMC} = 2.3 \times 10^9 M_\odot$ ; James & Ivory 2011); the SMC is in the  $X = \frac{1}{128}$  bin ( $\mathcal{M}_{S,SMC} = 5.3 \times 10^8 M_\odot$ ; James & Ivory 2011); M31 is in the  $X = 2$  bin ( $\mathcal{M}_{S,M31} = 1.0 \times 10^{11} M_\odot$ ; Hammer et al. 2007) and M33 is in the  $X = \frac{1}{8}$  bin ( $\mathcal{M}_{S,M33} = 5.7 \times 10^9 M_\odot$ ; Verley et al. 2009).

### 5.2.4 Halo mass

$$Y \times M_{H,MW} \quad (2)$$

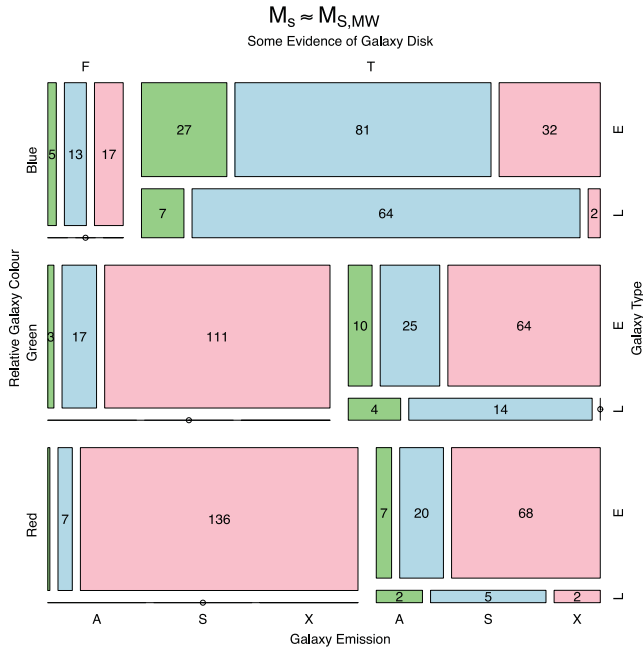
The discrete halo mass bins are defined in a similar vein to the stellar mass bins: adjacent bins vary in mass by a factor two, and each bin therefore contains a mass range of  $\pm 0.15$  dex. Halo masses are taken from R11, and are the variably scaled (with redshift and group multiplicity) dynamical masses that are discussed in that work.  $M_{H,MW}$  uses the halo mass of the MW calculated in Li & White (2008):  $M_{H,MW} = 2.5 \times 10^{12} M_\odot$ . The  $Y = 1$  bin corresponds to haloes with MW and M31 masses. The  $Y = 2$  bin corresponds to groups that have the same halo mass as the whole LG system ( $5.0 \times 10^{12} M_\odot$ ). Halo masses are 100 per cent complete for groups where  $\frac{1}{2} \leq Y \leq 2$ .

## 6 GALAXY PROPERTY DEPENDENCIES

### 6.1 The imperfect relationship between colour, emission, type and disc

Having established a set of contingencies that describe galaxies in a discrete manner, we can now construct mosaic plots that reveal the most significant dependencies. Combining information in this manner will give us an insight into what drives galaxy formation in galaxies with a stellar mass content similar to the MW. We can also determine which galaxy parameters are degenerate.

Galaxy populations are known to be bimodal in a number of properties (see Blanton et al. 2003; Baldry et al. 2004, 2006; Driver



**Figure 6.** Mosaic plot of Colour (Section 5.2.2)/Disc Presence (Section 5.1.2)/Type (Section 5.1.1)/Emission (Section 5.2.1). Subset used for this figure is  $\mathcal{M}_s = \mathcal{M}_{s,MW}$  (Section 5.2.3). Selection shown is the mass sample (Section 3). Observations: (i) Star formation predicts galaxy is blue and has a disc. (ii) No  $H\alpha$  emission predicts the galaxy is early-type.

et al. 2006; Taylor et al., in preparation). The main ones that we can investigate here are colour, disc presence, morphological type and line emission. A simplistic description of extragalactic astronomy might bracket galaxies into two main classes (or dependencies): blue, discy, star forming, late-type galaxies and red, discless, passive, early-type galaxies.

Fig. 6 shows the mosaic for Colour/Disc Presence/Type/Emission Class for galaxies where  $\mathcal{M}_s \sim \mathcal{M}_{s,MW}$  (within  $\pm 0.15$  dex of the MW stellar mass). While this mosaic plot contains a narrow range of galaxy stellar mass, it is immediately clear that there are no contingencies that unambiguously predict each other. For example, in this mass range, almost as large a fraction of green galaxies are non-star forming as red galaxies. This method of presenting bimodality is in contrast to more standard 2D scatter relations, with a third dimension simulated through successive cuts in the data (e.g. Blanton et al. 2003). With good clarity we are able to describe codependencies in four dimensions, while maintaining the relative counts of objects throughout in a graphically intuitive way by representing the number density with the size of the box. This latter point is important because this information is often lost in scatter plot contours. A final consideration is that we are able to display discrete and continuous characteristics on the same set of axes in a consistent manner.

Despite the complex dependencies we see in Fig. 6, there are two strong predictors in the data. Star-forming emission predicts the galaxy being blue and having a disc (but not its morphological type). Also, no emission predicts the galaxy being early type (but not the presence of a disc or colour, beyond being not blue). The reverse of these statements are much weaker predictors: blue galaxies with a disc are most likely to have star-forming emission, and early-type galaxies display a large mixture of emission classes, where 58 per cent have no emission. Colour is a strong predictor of disc presence (blue galaxies have a strong bias towards discs) but a much weaker

predictor of morphological type. This is mostly because early-type galaxies dominate for all colours. In general it is not accurate to assume any galaxy property inevitably correlates with another.

AGNs do not dominate any contingency, but we find they are most common for blue, early-type galaxies with evidence of a disc. Colour is a good AGN discriminator – they are much more likely ( $\sim 4:1$ ) in blue galaxies than red. Also, AGNs are 5 times more common in early-type galaxies compared to late-type galaxies. The strongest predictor though is disc presence. AGNs are six times more likely to reside in an  $L^*$  galaxy with a disc than one without. They are effectively non-existent in red, discless, early-type galaxies, so truly ‘dead’ systems will also be devoid of AGN  $H\alpha$  activity in the stellar mass range considered.

In the following sections we will have to be careful to extract the most significant of these 4 parameters when creating four-way mosaics that contain other variables. The most interesting (predictive or predicted) of these four contingencies vary case by case, and obviously depend on the physics being assessed and the time-scales being considered.

## 6.2 The group environment of $L^*$ Galaxies

The group environment plays a key role in the evolution of galaxies. In this section we investigate the effect that being in a group has on galaxies of a given stellar mass.

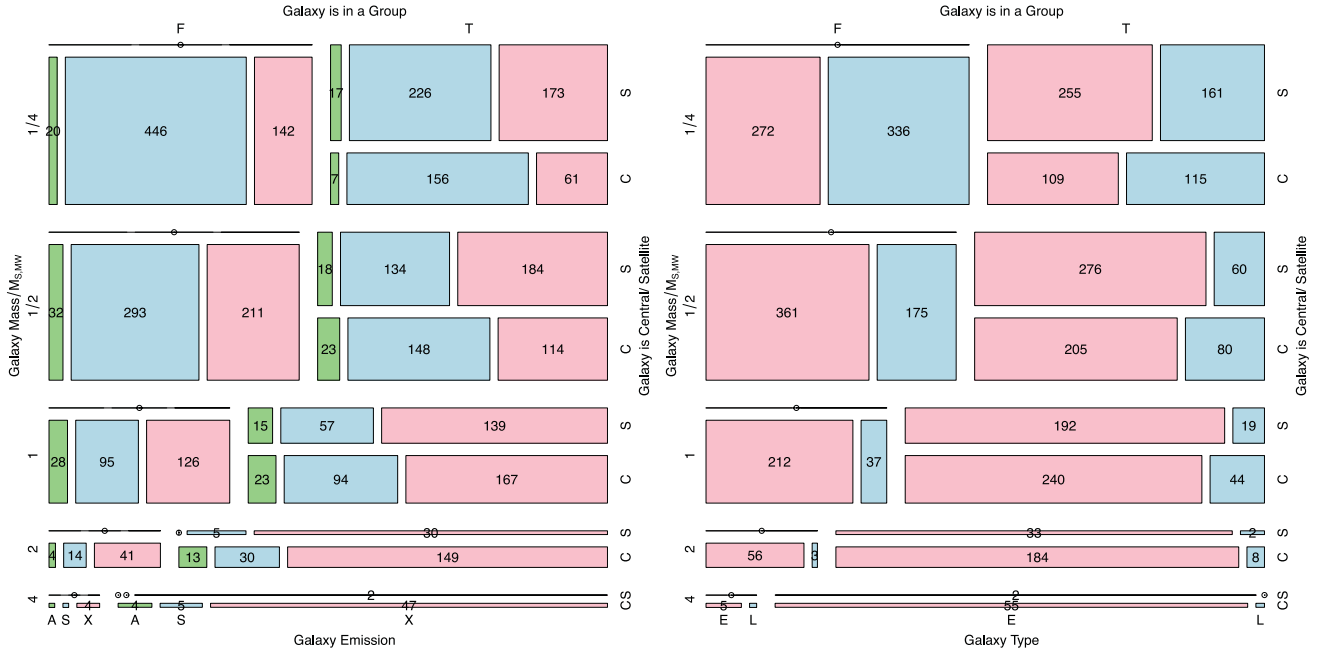
Fig. 7 shows two mosaic plots that only vary on the final contingency. The first three are stellar mass (down to our stellar mass limit of  $\mathcal{M}_s = \mathcal{M}_{s,MW}/4$ ), whether the galaxy is grouped and whether the galaxy is a central or satellite. The last contingency shown is emission class for the left-hand panel and morphological type for the right-hand panel. In these plots, we are able to observe down to LMC luminosity galaxies, so the term ‘ungrouped’ implies there are no near-by galaxies which have an  $r$ -band flux brighter than the LMC.

Some simple conclusions can be drawn immediately. Larger stellar mass galaxies are more likely to be grouped, and also more likely to be a central galaxy if they are in a group. Also, a galaxy is more likely to possess no star formation if its stellar mass is larger regardless of whether it is grouped.

More interesting is the clear role the group environment has on the emission class and morphological type of galaxies. Within grouped galaxies, and within the same stellar mass interval, there is a clear indication that satellite galaxies have a larger no-emission fraction than central galaxies. It should be emphasized that here we are comparing galaxies with the *same* intrinsic stellar mass, so we are not affected by biases regarding what varieties of galaxies tend to be grouped etc.

Central galaxies in groups have very similar emission class fractions to similar mass ungrouped galaxies, and satellite galaxies have significantly lower star-forming fractions. The morphological type fractions remain very similar for all masses between ungrouped galaxies and central galaxies, however satellite galaxies have a consistently larger early-type fraction. Galaxy morphological type shows much more dramatic stellar mass dependencies than the star formation fraction. For all contingencies, a much smaller fraction of galaxies are late type than star forming for the mass range investigated, and the drop with stellar mass is much more dramatic: 60 per cent to 9 per cent star forming between  $\frac{1}{4}\mathcal{M}_{s,MW}$  and  $4\mathcal{M}_{s,MW}$ , and 45 per cent to 2 per cent late type over the same mass range.

Fig. 8, which shows the raw population fractions as a function of stellar mass. Most fraction gradients trace each other quite closely.



**Figure 7.** Left-hand panel is a mosaic plot of stellar mass (Section 5.2.3)/galaxy in group (Section 5.1.3)/galaxy is central or satellite (Section 5.1.4)/emission (Section 5.2.1). Right-hand panel is a mosaic plot of stellar mass (Section 5.2.3)/galaxy in group (Section 5.1.3)/galaxy is central or satellite (Section 5.1.4)/type (Section 5.1.1). Ungrouped galaxies are referred to us ‘central’ since they are the largest galaxy in their multiplicity one group. Selection shown is the volume sample (see Section 3). Observations (left): (i) Larger stellar mass galaxies are more likely to be in a group. (ii) The fraction of star-forming satellite galaxies is less than observed for central galaxies of the same stellar mass. (iii) Central galaxies have similar emission fractions to ungrouped galaxies, which are in reality extremely stellar mass dominant central galaxies in their haloes. Observations (right): (iv) The fraction of late-type satellite galaxies is less than observed for central galaxies of the same stellar mass. (v) The fraction of late-type galaxies decreases more rapidly with stellar mass than the fraction of star-forming galaxies.

The notable exception is the disc fraction. This has a notably flatter dependency with stellar mass, and we see a pronounced break in the population fraction for grouped central galaxies. This population has a disc fraction that stays quite level at  $\sim 60$  per cent until the stellar mass is equal to  $M_{S,MW}$ , after which it drops quite steeply. A possible interpretation is that the most massive central galaxies are less able to accrete material (through either minor mergers or disc-like star formation triggered by gas accretion), causing the optical discs to become less common.

In general, satellite galaxies experience suppression of all features plotted in Fig. 8. The ungrouped galaxies look much more like grouped central galaxies. This is consistent with them being central galaxies of groups with undetectably large magnitude gaps, i.e. the second-rank member is less massive than the LMC (the stellar mass limit of the sample investigated). The main deviation is a significant disc fraction for more massive galaxies compared to the grouped BGGs, a feature consistent with ongoing minor merging. The grouped central galaxies will be more likely to be undergoing major interactions with more massive satellite galaxies, and such events can suppress evidence of any disc-like structure.

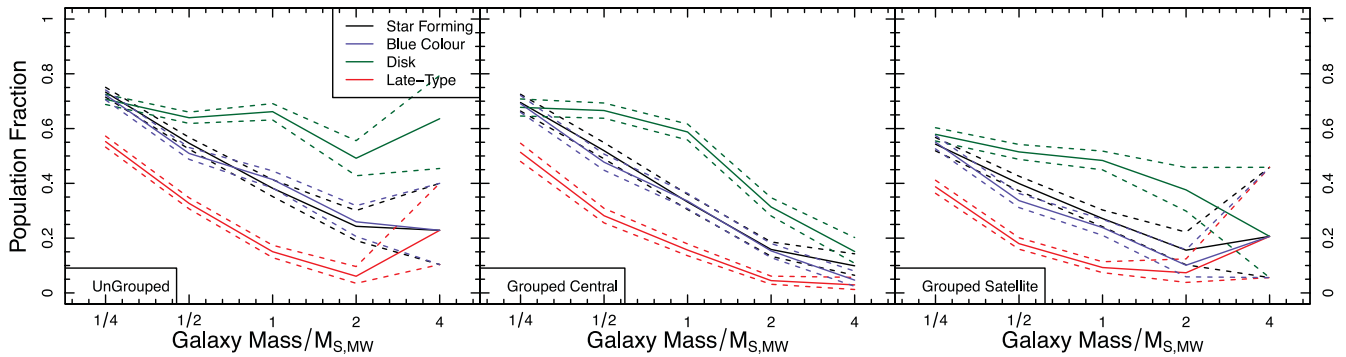
Fig. 9 allows us to extend the investigation into more detail, adding halo mass relative to the mass of the MW halo as a contingency and removing the grouped logic, i.e. only grouped galaxies are shown in the mosaic. We find that halo mass has a very weak effect on the star formation class of the central galaxy, but there is a tendency for more star-forming galaxies in less massive haloes. The satellites have a strong variation that depends on halo mass: for the same stellar mass galaxy the star-forming fraction is suppressed by a factor  $\sim 2$  between  $M = M_{H,MW}/4$  and  $M = 32M_{H,MW}$ . The strength of this suppression is strongest for the least massive

galaxies in the sample. This is in line with the findings of Grootes et al. (2012), who found that only the SFR of lower mass (likely to be satellite) galaxies showed a systematic decrease with halo mass. The morphological type is driven in a very similar manner (but the mosaic is not shown here), where the late-type fraction increases in line with the ‘S’ class emission shown in Fig. 9. AGN activity shows no statistically significant relationship with any group related property.

Combining this information with the earlier mosaics allows us to build a picture of galaxy evolution drivers for galaxies with similar stellar mass content to the MW. Merely being the central galaxy in a group of any mass provides a clear boost to the star-forming fraction of galaxies, but beyond this the halo mass plays no significant role, as shown by the lack of clear systematic effects with halo mass in Fig. 9.

The smallest satellite galaxies ( $M_{S,MW}/4$ ) do show a systematic effect, exhibiting suppression of the star-forming fraction as a function of halo mass. Since we have isolated out the effects of stellar mass and halo mass in the mosaics, this indicates that a galaxy’s relative stellar mass compared to other galaxies in its shared halo (which is the discriminator between central and satellite galaxies in this work) has an important role in its evolutionary path. Otherwise we would expect the emission fractions for similar stellar mass and similar halo mass galaxies (BGGs and satellites) to be closer, where the only missing descriptor is how dominant the mass is compared to the other galaxies in the group environment.

There are a number of possible explanations for such effects. One possibility is that satellite galaxies are modified during merger events, e.g. their gas is stripped upon entering a larger halo and their probability of forming stars is reduced accordingly. Another



**Figure 8.** Population fractions for different properties that are commonly associated with late-type populations: being morphologically late-type (by definition), star forming, having a disc and being blue in integrated colour. These are shown for ungrouped (left-hand panel), central (middle panel) and satellite (right-hand panel) galaxies. Dotted lines show the formal Bayesian error for a bimodal distribution, in this case the chance of galaxies being classed as stated in the legend, or not. Selection shown is the volume sample (see Section 3). Observations: (i) The disc fraction for grouped galaxies drops rapidly beyond  $\mathcal{M}_{S,MW}$ . (ii) Galaxy discs are more likely to be present in ungrouped galaxies for galaxies more massive than  $\mathcal{M}_{S,MW}$ , suggesting they can be easily disrupted by dynamical interactions.



**Figure 9.** Mosaic plot of galaxy is central or satellite (Section 5.1.4)/stellar mass (Section 5.2.3)/halo mass (Section 5.2.4)/emission (Section 5.2.1). The subset used for this figure is all grouped galaxies from the selection described in Section 3. There are no satellite galaxies with  $\mathcal{M}_s = 4\mathcal{M}_{S,MW}$ , so a line is drawn through these panels. Selection shown is the volume sample (see Section 3). Observations: (i) Satellites comprise a larger fraction of group stellar mass for more massive haloes. (ii) The least massive satellites ( $\mathcal{M}_{S,MW}/4$ ) experience a reduction in the star-forming fraction as a function of halo mass.

explanation, and one we investigate in detail later in this work, is that simple galaxy–galaxy interactions may be responsible for a large amount of this effect, i.e. the star-forming fraction of satellite galaxies is adjusted based on the probability it experiences different stellar mass galaxies. This latter effect will be a function of the stellar mass of the satellite and of the distribution of stellar masses present in the group. By our definition of central galaxy (most massive) there will be *at least* one galaxy more massive than any satellite galaxy in any group, whereas a central galaxy will be the most massive in all interactions. The effects of mass ratios on galaxy properties are investigated later in this paper.

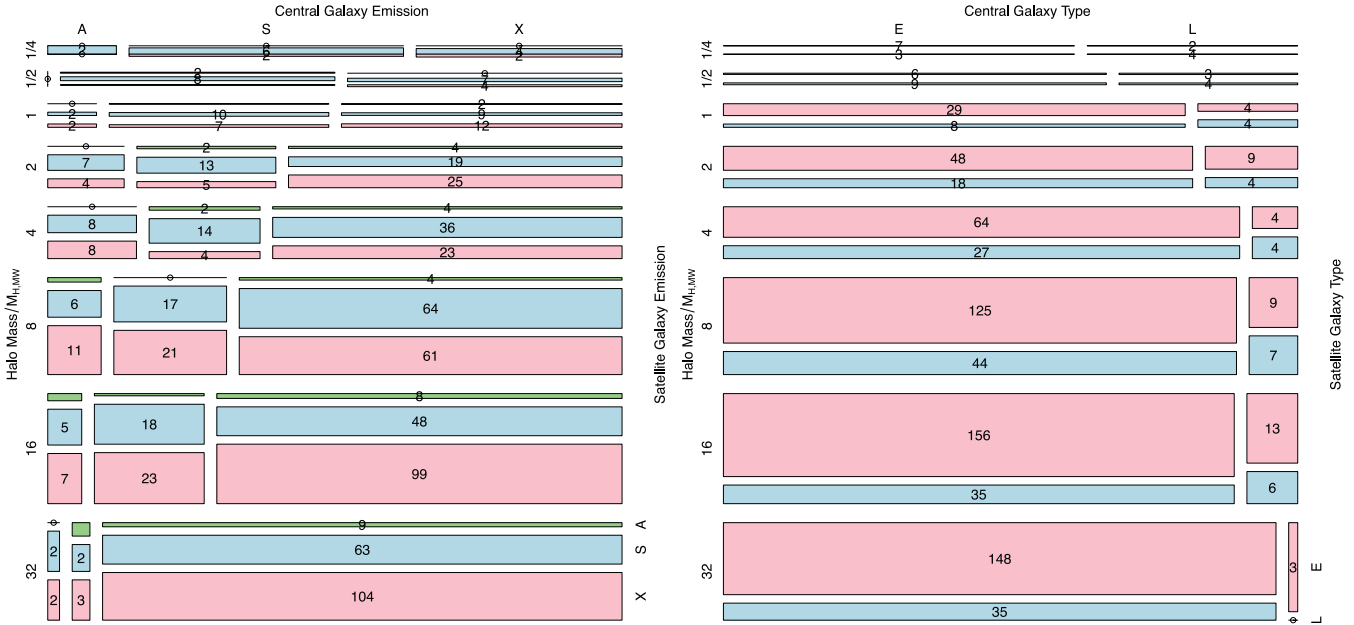
By definition the MW is found in a  $M_{H,MW}$  mass halo, and has a stellar mass of  $\mathcal{M}_{S,MW}$ . Since the subset explored is limited to the regime where we would expect to observe (and group) both the MW and the LMC, we can be confident the MW halo would be in the ‘grouped’ subset of data. The emission class of the MW is likely to be ‘S’, i.e. it has a small but observable amount of star formation occurring. Using these classifications, we can use information in Figs 7 and 9 to draw conclusions about the typicality of the MW compared to similar stellar mass galaxies. A similar stellar mass galaxy is more likely to be in a group than not (like the MW) and more likely to be the central galaxy than not (like the MW). Overall such a galaxy is more likely to possess no star formation than some (unlike the MW), and tends to be found in groups with halo masses slightly larger than the MW halo (the MW is in a slightly under-massive halo given its stellar mass).

### 6.2.1 Codependence of galaxy properties in groups

Galaxy conformity is potentially an important phenomenon. The effect was first highlighted in Weinmann et al. (2006), and reproduced in a larger selection of SDSS DR7 data using a different analysis in Kauffmann et al. (2013). The premise is that the properties of central and satellite galaxies are correlated within a halo. This is something we can investigate in using our GAMA data.

Fig. 10 demonstrates the degree to which we observe both emission class (LHS) and morphological type (RHS) conformity. What is clear is that in general we do recover compelling evidence for morphological type conformity (RHS), which is very much in line with the variety discussed in Weinmann et al. (2006), although in that case the ‘type’ was defined using a mixture of galaxy and sSFR (so it was not a purely visual classification scheme, like ours). The emission class demonstrates a weak amount of conformity (LHS), and is certainly not as compelling as the morphological conformity.

Later we investigate the conformity between galaxies in closely interacting pairs. The argument we form is that the conformity we observe for groups reflects the net result of many close interactions, hence the presence of strong morphological type conformity (it takes a long time for the orbits of stars to change radically enough to alter the morphological type) and the weak presence of emission class conformity (this is a fairly instantaneous response to interaction stimulation). This interpretation of the different varieties of conformity that we observe is discussed in detail below.



**Figure 10.** Left-hand panel if a mosaic plot of halo mass of satellite galaxy (Section 5.2.4 and 5.1.4)/emission of central galaxy (Section 5.2.1 and 5.1.4)/emission of satellite galaxy (Section 5.2.1 and 5.1.4). Left-hand panel if a mosaic plot of halo mass of satellite galaxy (Section 5.2.4 and 5.1.4)/type of central galaxy (Section 5.1.1 and 5.1.4)/type of satellite galaxy (Section 5.1.1 and 5.1.4) selection shown is the volume sample (see Section 3). Observations (left): (i) There is a weak tendency for star formation in a central galaxy to be accompanied by star formation in a satellite galaxy for all halo masses. (ii) AGN activity in the central galaxy does not predict any particular response in the satellite galaxy. Observations (left): (iii) There is a stronger tendency for morphological type in a central galaxy to be correlated with morphological type in the accompanying satellite galaxy for all but the largest halo masses. (iv) This result is in line with the suggestion of galaxy conformity in Weinmann et al. (2006) and Kauffmann et al. (2013).

### 6.2.2 SFRs in groups

Fig. 11 shows the stellar mass versus sSFR for all galaxies that are found in groups within 0.3 dex of the mass of the MW halo. Fig. 12 the stellar mass versus projected separation to the BGG for all satellites for the same subset of data (right-hand panel). For comparative purposes the median mass and sSFR are shown for different subsets, allowing us to see where the MW–Magellanic system lies.

Considering Fig. 11 first, it is clear that the MW is in an unusual part of parameter space considering it is a late-type galaxy: all nearby BGGs in stellar mass sSFR space are early-type galaxies. Globally the sSFR is entirely typical, but considering its morphological type (late-type) it is extremely low even allowing for the  $\pm 0.15$  dex in SFR error calculated in Robitaille & Whitney (2010). The MW is  $\sim 2.5$  times more massive than typical for the global population given its halo mass, and it is  $\sim 7$  times more massive than the typical late-type BGG in this subset. The LMC is very close to typical in terms of both mass and sSFR compared to the global and the late-type population. The SMC mass galaxies are not complete in the redshift range plotted, so conclusions cannot be drawn. Overall the MW–LMC system appears to be reasonably typical compared to detected  $G^3$ Cv1 groups if we ignore the fact that the BGG is late-type. Given its other properties, the fact that the MW is late-type makes it appear quite atypical. This finding is in line with the conclusions of Robotham et al. (2012).

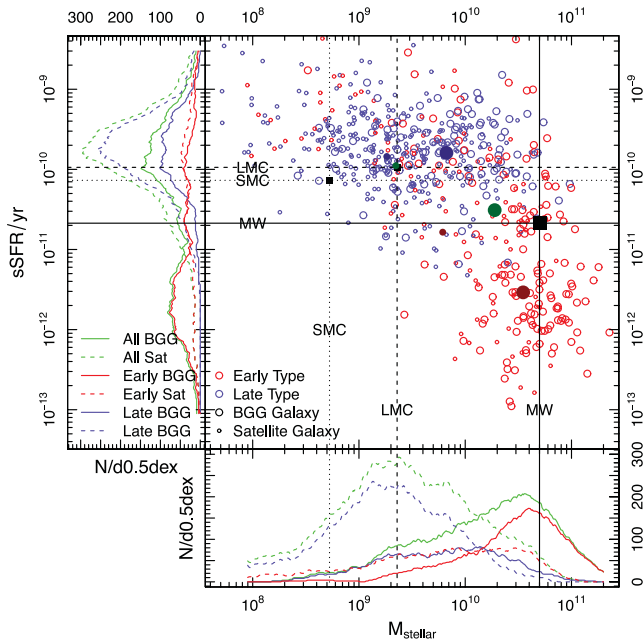
This analysis has to be taken with the caveat that *any* group catalogue will have a detection bias towards brighter galaxies being grouped. This is easy to imagine: if only the BGG was bright enough to be assigned a redshift and the second rank galaxy was just below the  $r$ -band survey limit than the BGG would not be assigned to a group. If the second rank galaxy was brighter (bringing it into

our selection limits) then we might be able to assign both galaxies to the same group. The strongest conclusion we can make in this work is that there is at least *some* MW halo mass parameter space that is occupied by  $\mathcal{M}_{S,MW}$  mass BGGs and  $\mathcal{M}_{S,LMC}$  mass satellites. This has important implications for galaxy occupation statistic work (e.g. Boylan-Kolchin et al. 2009, 2010, 2011; James & Ivory 2011; Tollerud et al. 2011; Weisz et al. 2011; Lovell et al. 2012; Robotham et al. 2012), since in principle we are biased towards selecting groups which look like our own MW halo, i.e. groups where the second rank galaxy (or brightest satellite galaxy) is fairly bright. The typicality of bright companion galaxies around MW stellar mass galaxies was investigated in detail in Robotham et al. (2012), with the conclusion that  $\sim 11.9$  per cent of MW stellar mass galaxies have an LMC stellar mass close pair. This was in broad agreement with the simulated predictions of Boylan-Kolchin et al. (2010).

In Fig. 12 we can see where the LMC and SMC sit in relation to satellites in similar mass groups when comparing stellar mass and separation to the BGG. Concentrating on the LMC, we can see that it lies closer to its BGG (the MW) than the median satellite, both compared to the global and the late-type populations. Early-type galaxies lie closer to the BGG than late-type galaxies, indicating that in an average sense the proximity of a satellite to the BGG affects its evolution. While the LMC might be far closer than the median, it is almost precisely at the *mode* of the mass–separation parameter space (shown by light-green contours in Fig. 12).

### 6.3 The pair environment of $L^*$ Galaxies

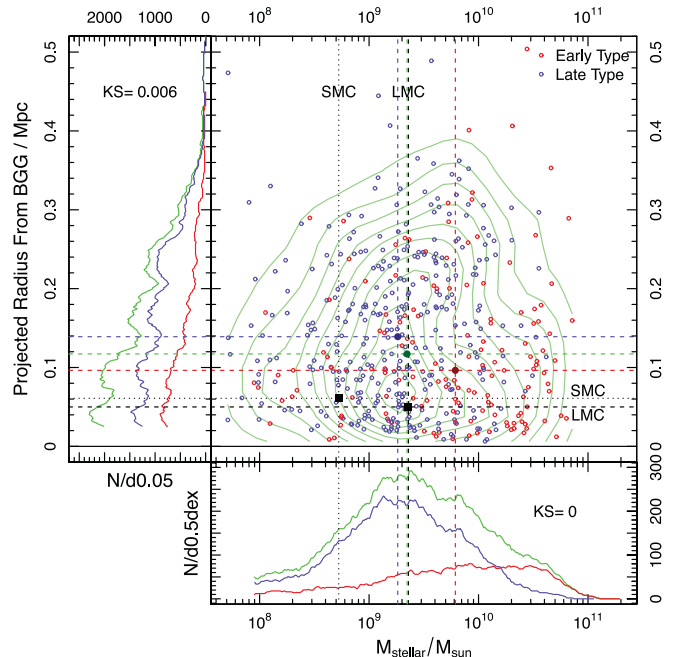
Fig. 13 shows the distribution of emission class in dynamical phase space for galaxies with a stellar mass within 0.3 dex of  $\mathcal{M}_{S,MW}$ . This plot was also assessed as a function of morphological type,



**Figure 11.** Stellar mass versus sSFR of all pairs where the group is within 0.3 dex of the halo mass of the MW (here assumed to be  $2.5 \times 10^{12} M_{\odot}$ ). The colouring refers to morphological type (blue = late-type, red-early-type), and size specifies whether the galaxy is the BGG (large circle) or satellite (small circle). The location of the MW/LMC/SMC is over-plotted for reference. The population medians are over-plotted, where blue filled points are for late-type, red filled points are for early-type, and green filled points are the full population. For the MW, the stellar mass is from Flynn et al. (2006) and the SFR is from Robitaille & Whitney (2010). SMC and LMC stellar masses and SFRs are both from James & Ivory (2011). The density of counts is projected on to the  $x$ - and  $y$ -axes, and is separated into BGG / satellite and early / late / combined populations. This plot allows us to see where the MW system sits relative to similar halo mass groups. Selection shown is the volume sample (see Section 3). Observations: (i) The MW is close to the median sSFR and stellar mass for all central galaxies in MW halo mass groups, but is marginally more massive than typical and has a smaller sSFR (position relative to green point). (ii) Adding in the prior knowledge that the MW is late-type, it has an extremely low sSFR and is much more massive than typical for central galaxies in MW halo mass groups. (iii) The LMC is almost exactly at the median position expected for late-type satellite galaxies in MW halo mass groups. The caveat here is that to qualify as a group biases, the group sample to include groups with satellites as bright as the LMC (since this is the depth limit).

disc presence and colour. In all cases there are no strong population gradients within the dynamical phase space explored in this work as either a function of the projected separation or the radial velocity separation. There are clear offsets between the fractions of different classes and due to a galaxy being the major or minor partner in the galaxy pair system (this is discussed in detail later).

There is also an increased likelihood that MW stellar mass minor pair galaxies will be passive at small projected separations, compared to if they are the major pair galaxy. However, in general the constancy of the star-forming fraction in these distributions indicate we will not be strongly biased by our somewhat arbitrary definition of a galaxy pair in our analysis, and we will limit the discussion to the full dynamical selection criteria for pairs (rather than refining discussion to sub-bins of Fig. 13).



**Figure 12.** Stellar mass versus projected separation from the BGG for all satellites. The LMC and SMC are over-plotted for comparison. The solid circles show the medians for different subsets, where green, blue and red refers to all, late-type and early-type, and large and small refer to BGG and satellite. The light-green contours show the local density of the full sample. SMC and LMC stellar masses are both from James & Ivory (2011). The separation between the MW and LMC/SMC is taken from Nichols et al. (2011). The density of counts is projected on to the  $x$ - and  $y$ -axes, and is separated into early/late-type populations. The density plots include a KS test  $p$ -value comparison of the early and late-type distributions. Selection shown is the volume sample (see Section 3). Observations: (i) The LMC is much closer to its central group galaxy (the MW) than is typical for satellites in MW mass groups. (ii) The KS test comparison of early-type and late-type stellar masses/projected separations for satellites shows it is extremely unlikely the two distributions are sampled from the same parent sample. (iii) Early-type satellites tend to be  $\sim 2/3$  the projected separation to the BGG compared to late-type galaxies.

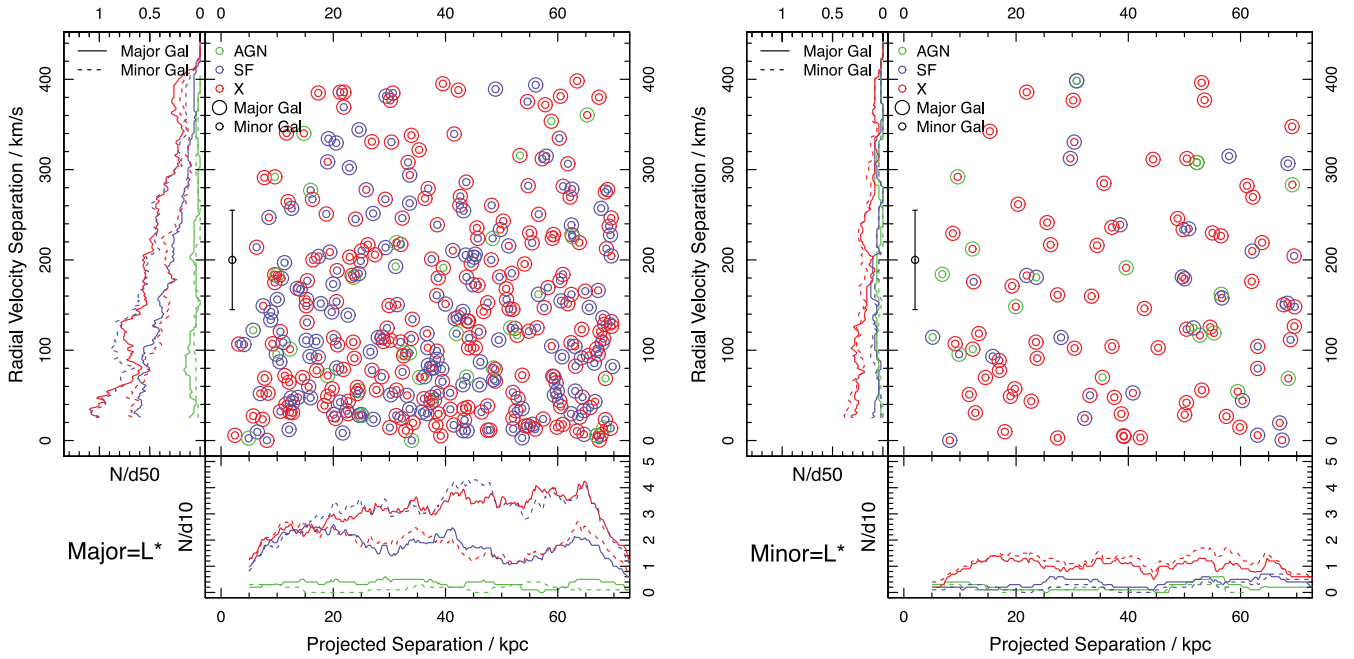
### 6.3.1 Pair fractions by mass

Having considered the role of the group environment on a given stellar mass galaxy, we can now investigate the effects galaxy pairs have on the evolution of the pair galaxies. This will probe the MW–LMC pair system directly, and allow us to determine how unusual this system is. The pair catalogue used for this work is discussed in Section 2.4. Due to the redshift, spatial and velocity separation selections we expect all MW–LMC like systems to be observed and complete (i.e. for  $z < 0.089$ ,  $v_{\text{sep-rad}} < 400 \text{ km s}^{-1}$  and  $r_{\text{sep-proj}} < 70 \text{ kpc}$ ).

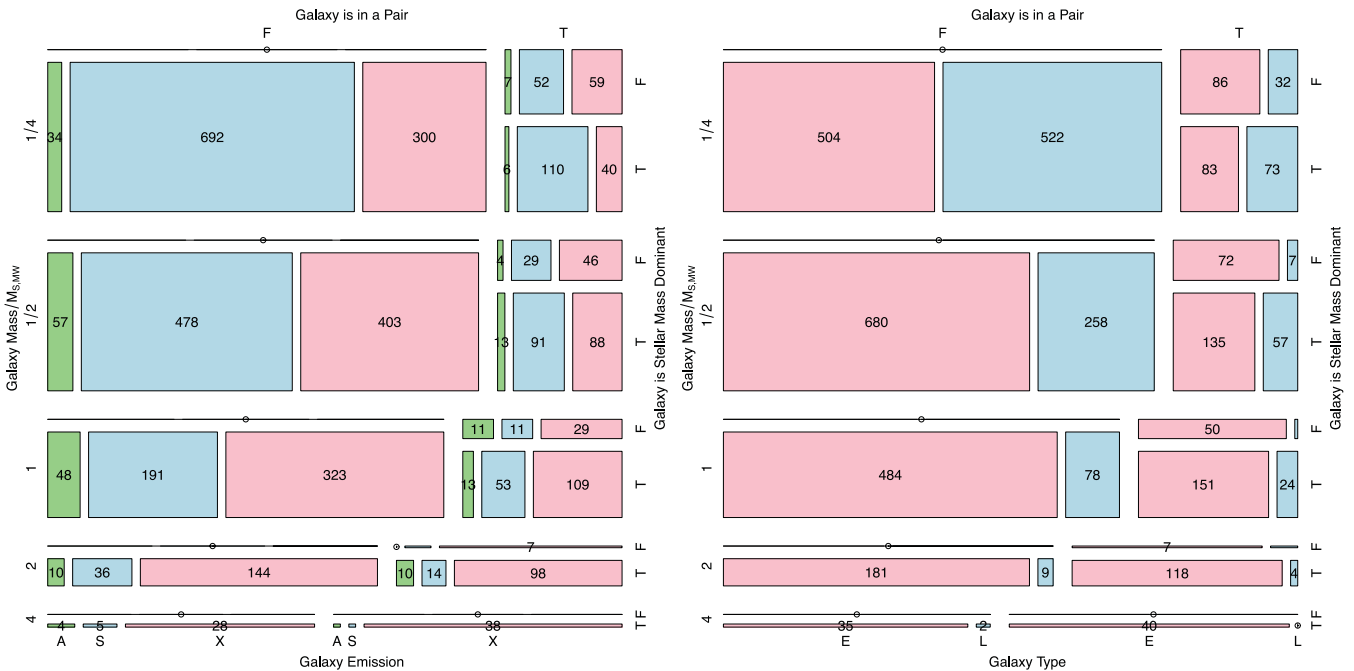
The left-hand panel of Fig. 14 shows the mosaic for Galaxy stellar mass/whether the galaxy is in a pair/whether the galaxy is the largest in the pair/ emission class. The right-hand panel exchanges the emission class contingency for morphological type.

Immediate observations regarding the left-hand panel of Fig. 14 are that more massive galaxies are more likely to be found in a pair meeting our selection criteria, but in all cases the majority of galaxies are not in pairs. The non-paired emission class distributions are most similar to paired galaxies when the galaxy is more massive. This makes sense because we are only confident the galaxy is not paired down to a minor mass limit of the LMC, so it might be





**Figure 13.** Projected separation versus velocity separation between galaxy pairs, where the major/minor galaxy is within 0.3 dex of stellar mass of the MW (left/right). Galaxy morphological type is indicated by colour, and pair designation by point type. Typical velocity errors (55 km/s) are shown on the LHS of the plot windows. Selection shown is the volume sample (see Section 3). Observations (left): (i) A much larger fraction of  $\mathcal{M}_{S,MW}$  mass galaxies are the major galaxy in a pair. (ii) When  $\mathcal{M}_{S,MW}$  mass galaxies are the major pair galaxy, the star-forming–passive (blue–red lines) fraction is close to 1–1 at small projected separations. Observations (right): (iii) When  $\mathcal{M}_{S,MW}$  mass galaxies are the minor pair galaxy, they have a much reduced star forming/passive ratio (ratio of blue/red dashed lines) at small projected separations compared to when they are the major pair galaxy.



**Figure 14.** Left-hand panel is a mosaic plot of stellar mass (Section 5.2.3)/galaxy in pair (Section 5.1.5)/largest galaxy in pair (Section 5.1.6)/emission (Section 5.2.1). Right-hand panel is a mosaic plot of stellar mass (Section 5.2.3)/galaxy in pair (Section 5.1.5)/largest galaxy in pair (Section 5.1.6)/type (Section 5.1.1). Selection shown is the volume sample (see Section 3). Observations (left): (i) More massive galaxies are more likely to be in a pair. (ii) The star-forming fraction for major pair galaxies is similar to that observed for unpaired galaxies. (iii) Minor pair galaxies have a smaller star-forming fraction than major pair galaxies for the same stellar mass. Together with the above remark, this indicates that pairwise interactions have a net role of suppressing of star formation, rather than initiating it (though both effects are likely to occur). Observations (right): (iv) Minor pair galaxies have a smaller late-type fraction than major pair galaxies for the same stellar mass, and this dependency is more dramatic than that seen for star formation (above). (v) The late-type fraction drops more rapidly with stellar mass than the star-forming fraction, particularly for minor pair galaxies. (vi) The MW would be more typical if they were early type, given its stellar mass. This agrees with the distributions shown in Fig. 11.

the dominant member of a pair with an even less massive minor companion. There is a weak tendency for the minor galaxy in a pair to have a smaller star-forming fraction than it would if it was the largest galaxy in the pair, or if it was unpaired. This is a similar result to the effects discussed for satellite galaxies in Section 6.2.

The trends seen in the right-hand panel of Fig. 14 are similar, where the late-type populations follow the same population trends as the ‘S’ class emission galaxies discussed above. In general we see that morphological type is mostly driven by galaxy stellar mass. One stronger dependency noted is that late-type galaxies of a given stellar mass are suppressed if it is the minor component of a pair system relative to when it is the dominant component. The relative amount of suppression is more dramatic than that seen for star formation. When the stellar mass is  $\frac{1}{4}M_{S,MW}$ ,  $\sim 2$  times as many galaxies are dominant pair star-forming galaxies compared to minor pair star-forming galaxies, and the late-type ratios are similar. However, when the stellar mass is  $M_{S,MW}$  the ratio is  $\sim 5$  for the star-forming fractions, but 24 for the late-type fractions. The difference between these numbers is independent of stellar mass effects, and shows a real difference in response to close pairwise interactions between star formation and morphology. Perhaps surprisingly, it is morphology that is the best tracer of such events.

Since being in a galaxy pair is potentially a very brief state given our pair selection requirements, this suggests that morphology is actually sensitive to dynamic events on short time-scales, but displays the effects for longer periods of time compared to the star-forming fraction. That is, there is a higher chance of observing modified morphology due to pairwise interaction than modified star formation.

Fig. 15 is similar to Fig. 8, but is now plotted for galaxies in pairs (rather than groups) and major or minor pair component (rather than group central or satellite galaxy). Galaxy colour displays the most consistency across the three selection classes plotted, while disc presence, star formation fraction and late-type fraction show a strong variation between the unpaired and minor pair galaxy classes for a given amount of galaxy stellar mass. A dramatic feature is that the most massive minor pair galaxies show a statistically significant increase in the disc fraction relative to the unpaired case. This sample was reanalysed, and 14/16 of minor pair galaxies with  $M_s \geq M_{S,MW}$  show extremely dramatic and asymmetrical spiral features associated with the pairwise interaction. This is in contrast to similar stellar mass systems with discs in the rest of the sample, these tending to be smooth symmetrical discs. Since in these ex-

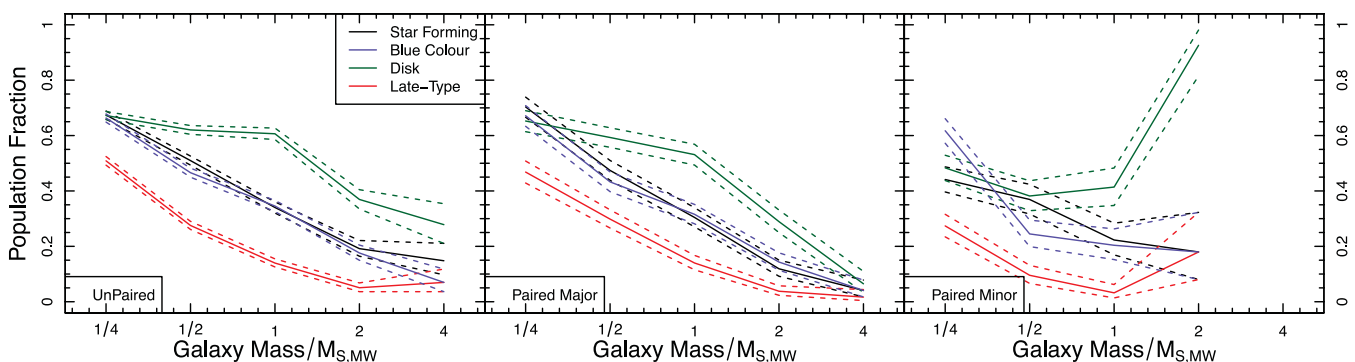
treme cases the dominant galaxy would have to be extremely large, the disc observed appears to be one generated by near-neighbour tidal disruption processes rather than gentle accretion. Such structures are transient (on the order of  $\sim 5$  Gyr) since it is the brief close passage between galaxies that generates them.

The grouped central galaxies portray much the same correlations as the dominant paired galaxies. This can be partly understood by the large fraction of population overlap: 63 per cent of dominant (or major) galaxy pairs are also the central galaxy of a galaxy group in the sample investigated. Equally, the ungrouped galaxies share similar population fraction trends to the unpaired galaxies. The notable exception is the galaxy disc fraction – there is a continuous downwards trend with stellar mass for the unpaired galaxies, while for the ungrouped galaxies we see evidence for a flattened correlation. Considering dynamic time-scales, it is likely an ungrouped galaxy has experienced a longer delay since the last major interaction than an unpaired galaxy, since not being grouped is a stronger isolation criterion than not being paired (a larger fraction of galaxies are unpaired compared to ungrouped). This fits with the hypothesis that that presence of a disc typically suggests a longer period of time since a major galaxy–galaxy interaction has ended, since these should disrupt the visible accretion-type disc (Barnes & Hernquist 1992a). The exception is complicated extended disc-like structures observed *during* tidal interactions between massive galaxies. These are formed by brief and violent close passages (see early discussion of such systems in Kormendy 1977).

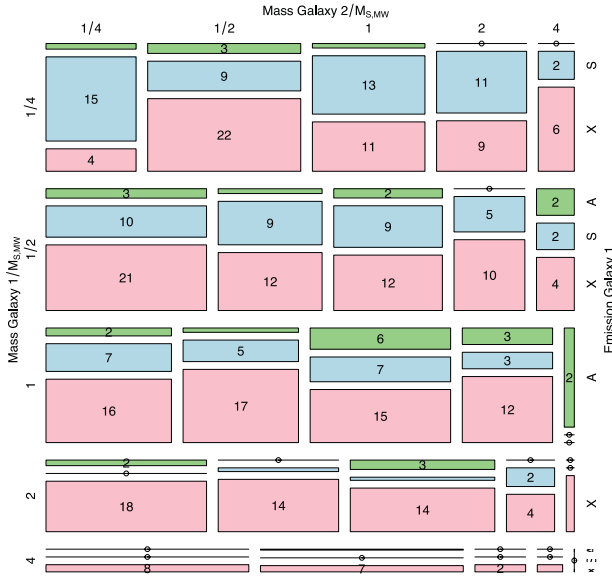
In Section 6.2 we found that the lowest Star-forming fractions were found in grouped satellite galaxies. An equivalent statement cannot be made regarding pairs: the lowest star-forming fractions are found for paired galaxies where the galaxy is the lowest mass member in the pair. In the next section we investigate how star formation can be suppressed and enhanced in pairs in more detail.

### 6.3.2 Effects of galaxy stellar mass ratios in pairs

A galaxy that is in a pair can be in a number of different mass ratio combinations, i.e. a given stellar mass galaxy can either be the larger (dominant or major) or smaller (minor) within a pair. The complication of this is that lower stellar mass galaxies near our survey limit will be incomplete since there will not be any lower mass galaxies within our sample selection for them to be paired with. These galaxies will therefore only appear as the minor partner in a galaxy pair.



**Figure 15.** Population fractions for different properties that are commonly associated with late-type populations: being morphologically late-type (by definition), star forming, having a disc and being blue in integrated colour. These are shown for unpaired (left-hand panel), dominant galaxy in pair (middle) and minor galaxy in pair (right) galaxies. Dotted lines show the formal Bayesian error for a bimodal distribution, in this case the chance of galaxies being classed as stated in the legend, or not. Selection shown is the volume sample (see Section 3). Observations: (i) The most massive minor pair galaxies show a statistically significant increase in the disc fraction relative to the unpaired case.



**Figure 16.** Mosaic plot of stellar mass of galaxy 1 (Section 5.2.3)/stellar mass of galaxy 2 (Section 5.2.3)/emission of galaxy 1 (Section 5.2.1). Selection shown is the volume sample (see Section 3). Observations: (i) Galaxy–galaxy pairs with similar stellar masses have the largest star-forming fraction, though this relationship is noisy. (ii) The star-forming fraction is driven most strongly by stellar mass; pairwise interactions have a weaker effect.

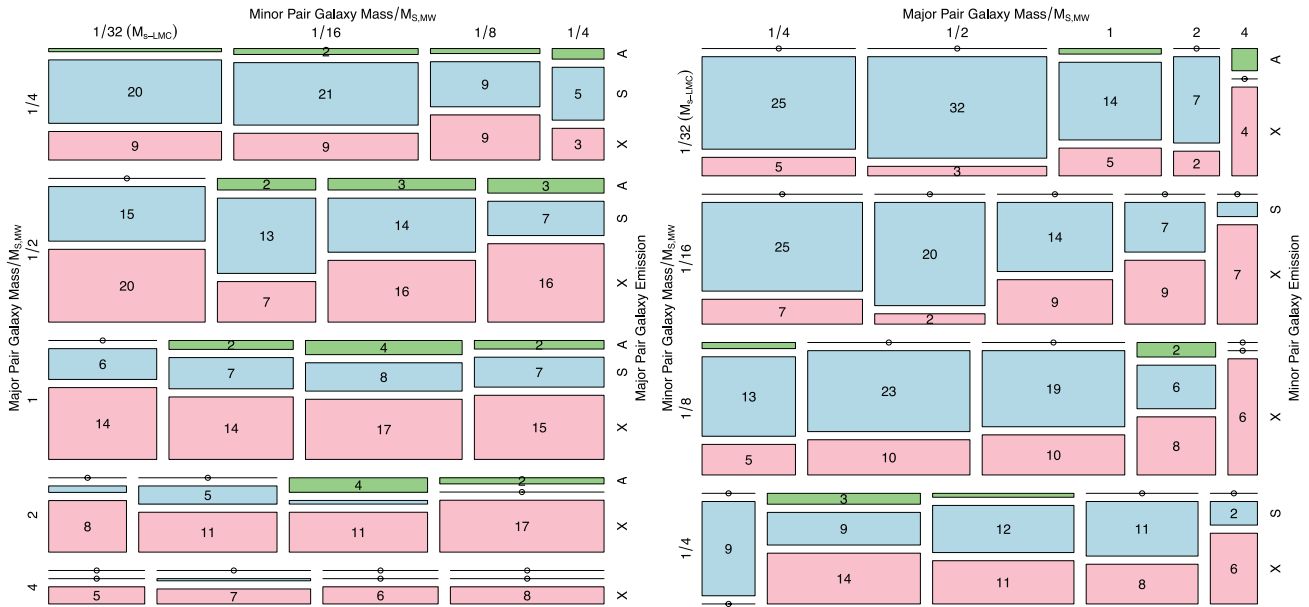
With this effect in mind, Fig. 16 shows the relationship for galaxy pairs in a narrow mass ratio regime where both the lower mass and higher mass galaxies are within our selection criteria:  $\mathcal{M}_{S,MW}/4 < \mathcal{M}_s < 4\mathcal{M}_{S,MW}$ .

The grids in Fig. 16 are aligned to reveal how the emission class of a galaxy varies as a function of the other galaxy pair mass. The number of galaxies within diagonally mirrored emission class cells will add up to the same value, but they will be distributed differently since the emission class is always shown for galaxy 1 (i.e. they correspond to the masses plotted on the left). There is a preference for galaxy pairs with similar masses to have the most evidence for star formation. When the secondary galaxy is much more or less massive the star-forming fraction drops appreciably. This implies that major merger (i.e. equal mass) interactions have the most significant star formation enhancing effect in galaxy pairs, where unpaired galaxies have as high (or higher) star-forming fractions.

To investigate how these trends continue beyond the stellar mass range considered above, we can explore a larger sample where we force galaxy 1 to always be the larger galaxy in the pair and within the volume and stellar mass limits ( $\mathcal{M}_{S,MW}/4 < \mathcal{M}_s < 4\mathcal{M}_{S,MW}$ ). Galaxy 2 has to be more massive than the LMC and less massive than the least massive galaxies in the galaxy 2 sample in Fig. 16 to ensure we are only considering a volume limited part of parameter space (i.e.  $\mathcal{M}_{S,MW}/32(\sim \mathcal{M}_{S,LMC}) < \mathcal{M}_s < \mathcal{M}_{S,MW}/4$ ).

The result of this selection is shown in Fig. 17, where the first two contingencies are the two galaxy masses in the pair and the third contingency is the emission class of the more massive galaxy (left-hand panel) and the emission class of the less massive galaxy (right-hand panel).

Considering the left-hand panel of Fig. 17 first, we can continue the trends observed in Fig. 16. There is weak evidence that while the star-forming fraction peaks near 1–1 mass ratio pairs, it then plateaus as the lower mass companion becomes less massive. The average star-forming fractions of the dominant galaxy are similar to unpaired galaxies (see the left-hand panel of Fig. 14), so overall they



**Figure 17.** Left-hand panel is a mosaic plot of stellar mass of the major pair galaxy (Sections 5.2.3 and 5.1.6)/stellar mass of the minor pair galaxy (Section 5.2.3 and 5.1.6)/emission of the major pair galaxy (Sections 5.2.1 and 5.1.6). Right-hand panel is a mosaic plot of stellar mass of the minor pair galaxy (Section 5.2.3 and 5.1.6)/stellar mass of the major pair galaxy (Section 5.2.3 and 5.1.6)/emission of the minor pair galaxy (Sections 5.2.1 and 5.1.6). Selection shown is the volume sample (see Section 3). Observations (left): (i) The star-forming fraction of major galaxies is largely unaffected by having pair companions that are much less massive than itself. Star formation in the major pair galaxy of MW–LMC analogues is largely unaffected by the LMC mass minor companion. Observations (right): (i) Star-forming galaxies for minor pair galaxies show a decline as the major pair galaxy mass increase. (ii) Where there is population overlap with the left-hand panel, the minor galaxies possess a smaller star-forming fraction.

appear enhanced when mass ratio is 1–1, and suppressed otherwise. This finding requires more data to be robustly confirmed since the number statistics are small, but this work suggests galaxy–galaxy pair stellar mass ratios nearly always suppress star formation, with the net effect at a minimum when the galaxy masses are most similar. This is an interesting point, since gravitational braking is at its most efficient when the mass ratio is close to 1–1, i.e. these will be the galaxies most likely to be undergoing full mergers and not simply transitory interactions.

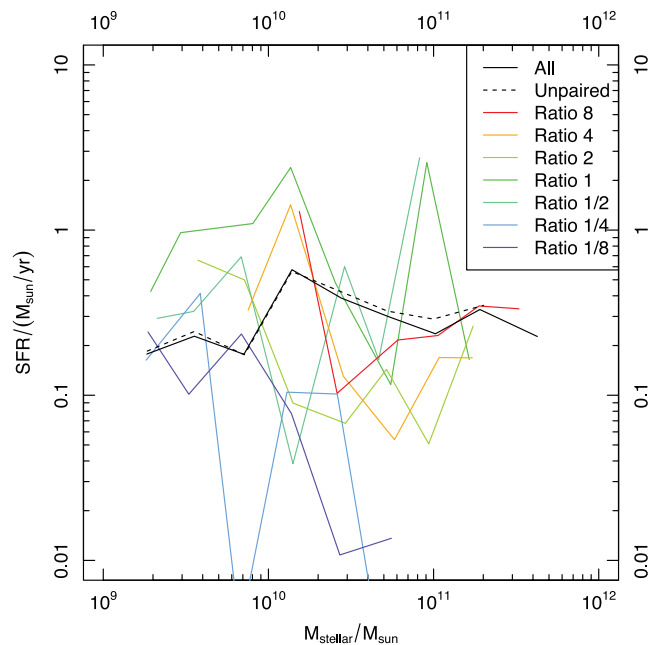
The right-hand panel of Fig. 17 flips round the first two contingencies of the left-hand panel so that the emission classes always correspond to the left side divisions. Therefore, it should not be immediately surprising that we see much larger star-forming fractions in the right-hand panel – it shows the grids for typically lower mass galaxies. A couple of observations can be drawn though.

The first is that the one mass range that exists for both minor and major pair galaxies is  $\mathcal{M}_s \leq \mathcal{M}_{S,MW}/4$  (the top strip of the left-hand panel and the bottom strip of the right-hand panel) and the star-forming fraction is clearly higher when it is the larger partner in the pair. This tallies with the observations regarding Fig. 14, where we saw that dominant pair galaxies of a given stellar mass have more ‘S’ class emission.

The second is that minor galaxy emission is more sensitive to the mass of the major pair galaxy than vice-versa. This is clear from the star-forming fractions that vary strongly as a function of major galaxy mass: the over-riding trend being the more dominant the mass of the major galaxy (the larger the mass ratio) the smaller the fraction of strongly emitting galaxies.

Going further, Fig. 18 shows the runnings median SFR as a function of stellar mass for different volume limited samples of galaxy–galaxy pair mass ratios. The data is of low S/N quality, but a clear feature is that the 1–1 mass ratio population (green solid line) is typically the population with the highest SFR given the stellar mass. So while the *fraction* of galaxies forming stars is suppressed in all galaxy pairs, the amount of star formation occurring in those galaxies that are still forming stars actually increases when the interaction mass ratio is close to 1–1. As a counter to this, the minor pair galaxies with the smallest mass ratios (bluer lines) are those with the least SFR given the stellar mass. The y-axis separation between the 8–1 mass ratio SFR (red line) and the 1/8–1 mass ratio SFR (blue line) is roughly 1 dex.

The combination of information from the figures presented in this section produce a compelling picture when it is all pieced together. The emission state of dominant pair galaxies is only significantly altered when the pair ratio is close to 1, otherwise we observe the same galaxies as we would in an unpaired system. The minor galaxy in a pair is greatly affected by its companion galaxy, and while all star formation is suppressed relative to equivalent stellar mass unpaired galaxies, it becomes lower still when the mass ratio increases (i.e. the major galaxy becomes more massive and dominant within the pair). The baryonic physics that fully explains these effects is beyond the scope of this work, but there are likely to be a number of physical processes competing for dominance during a galaxy–galaxy interaction, some of these will trigger star formation and some will suppress it. The *net* effect of the interactions analysed in this work is to suppress star formation, with the caveat that this is least evident for similar stellar mass galaxy pairs. While the fraction of galaxies forming stars is generally lower in a galaxy pair, the absolute peak of star formation seen is for a 1–1 mass ratio interaction.

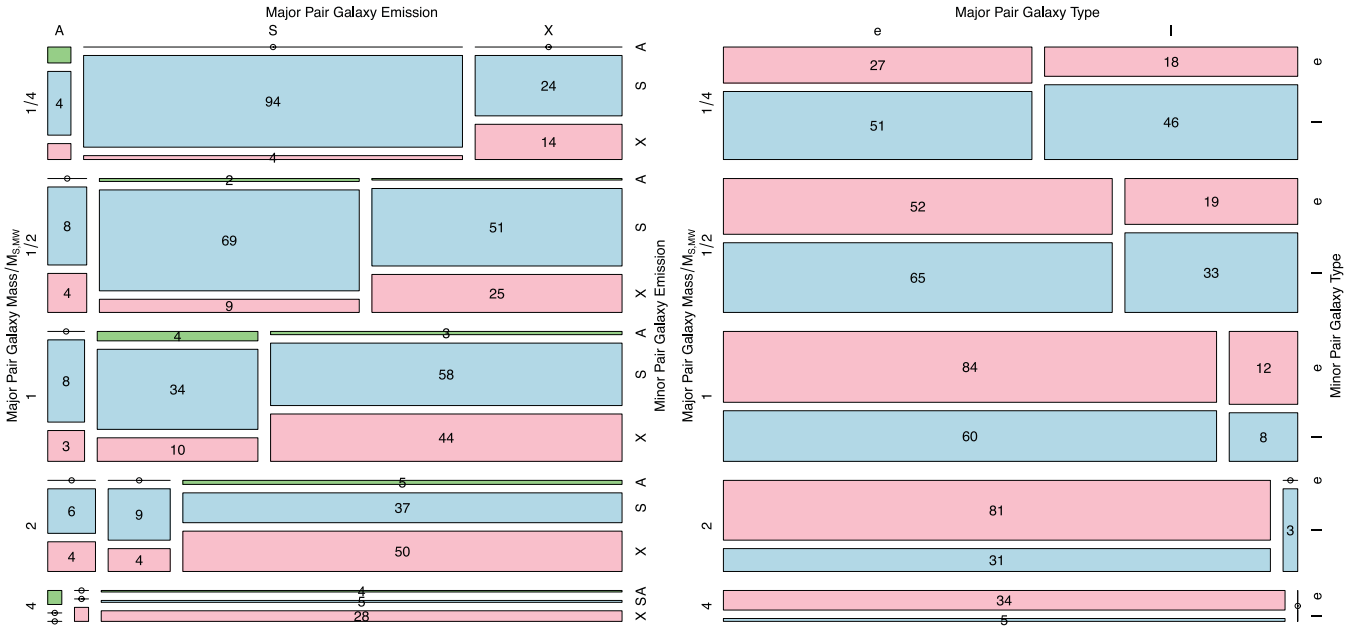


**Figure 18.** Stellar mass versus star formation rate (SFR) for volume limited samples of galaxy pairs with varying pair mass ratios. All samples are volume limited, with the stellar mass limits adjusted as a function of the pair ratio. Black solid line shows the running median for all galaxies in our volume limited sample. Black dashed line shows the running median for all unpaired galaxies (given our observational depth constraints). Different colours refer to different mass ratios, with 1–1 mass ratio pairs shown as a green line. Selection shown is the volume sample (see Section 3). Observations: (i) Separate to the star-forming fraction (shown in the mosaic plots), here we see that the actual amount of star formation is modified by the pair environment: mass ratios close to 1–1 (green line) typically display the highest median SFR for a given amount of stellar mass. (ii) Star formation is dramatically suppressed when the mass ratio is large and the galaxy under consideration is the minor pair galaxy (blue line).

### 6.3.3 Codependence of galaxy properties in pairs

Galaxies in pairs can have a number of emission state permutations, hence it is interesting to investigate whether the emission strength in one member of a galaxy in a pair predicts the other: i.e. if the larger galaxy in a pair is experiencing star formation, does it make it any more likely that the minor partner is also star forming?

Fig. 19 shows a mosaic plot of how the emission class of the dominant pair galaxy predicts the emission class of the minor pair galaxy for a given stellar mass of the dominant galaxy (LHS) or a similar plot showing the codependence of morphological type. The clear indication from this plot is that the emission classes are highly codependent. Star-forming emission in either the dominant or minor galaxy predicts it in the other, but this is particularly true for dominant predicting minor. Taking  $\mathcal{M}_{S,MW}$  stellar mass galaxies, if star formation is present in the dominant galaxy it is 71 per cent likely that the minor galaxy is also star forming. If the same stellar mass dominant galaxy instead showed no emission (‘X’ class) then it is only 55 per cent likely that the minor galaxy has star-forming emission. The strength of this conformity effect is remarkably regular for all stellar masses, although the fraction of star-forming dominant galaxies with stellar masses above  $\mathcal{M}_{S,MW}$  is very small.



**Figure 19.** Left-hand panel is a mosaic plot of stellar mass of the major pair galaxy (Sections 5.2.3 and 5.1.6)/emission of the major pair galaxy (Sections 5.2.3 and 5.1.6)/emission of the minor pair galaxy (Sections 5.2.1 and 5.1.6). Right-hand panel is a mosaic plot of stellar mass of the major pair galaxy (Sections 5.2.3 and 5.1.6)/type of the major pair galaxy (Sections 5.1.1 and 5.1.6)/type of the minor pair galaxy (Sections 5.1.1 and 5.1.6). Selection shown is the volume sample (see Section 3). Observations (left): (i) There is a clear tendency for star formation in a major pair galaxy to be accompanied by star formation in a minor pair galaxy for all major pair galaxy stellar masses. (ii) If the major pair galaxy does not show evidence for star formation, it is less likely that the minor pair galaxy will be forming stars. (iii) When the major pair galaxy shows evidence for AGN activity, the likelihood of star formation in the minor pair galaxy is between the above two extremes. Observations (right): (iv) There is weaker tendency for morphological type of a major pair galaxy to be correlated with the morphological type of the accompanying minor pair galaxy. (v) This is the pairwise equivalent of galaxy conformity, described for groups and isolated galaxy systems in Weinmann et al. (2006) and Kauffmann et al. (2013), respectively. (vi) The relative strengths of the galaxy conformity effect seen here are reversed compared to groups (above). There the morphological type showed stronger conformity, whereas when galaxy pairs are considered the conformity is greatest for emission class.

The ‘X’ class prediction is also very strong, and it becomes more dominant for higher mass dominant galaxies. For all masses, minor galaxy star formation is most likely when the dominant is star forming and least likely when it shows no signs of emission. If the dominant galaxy hosts an AGN then the chance of the minor galaxy showing star-forming emission is broadly between these two extremes. If we hypothesise that AGNs form after a period of star formation, this would suggest that whatever mechanism has stimulated an AGN in the dominant pair galaxy (and shut down star formation) is correlated with a small suppressive effect on the star formation properties of the nearby minor pair galaxies. This correlation does not suggest causation, it could just indicate that there is a natural temporal ordering associated with these observations.

The codependence of morphological type is less clear cut, with only a weak tendency for the morphological type of the major pair galaxy to predict the morphological type of the minor pair galaxy. This is the reverse of what we saw for grouped galaxies, where the morphological type conformity between central and satellite galaxies was very strong (in line with the observations of Weinmann et al. 2006).

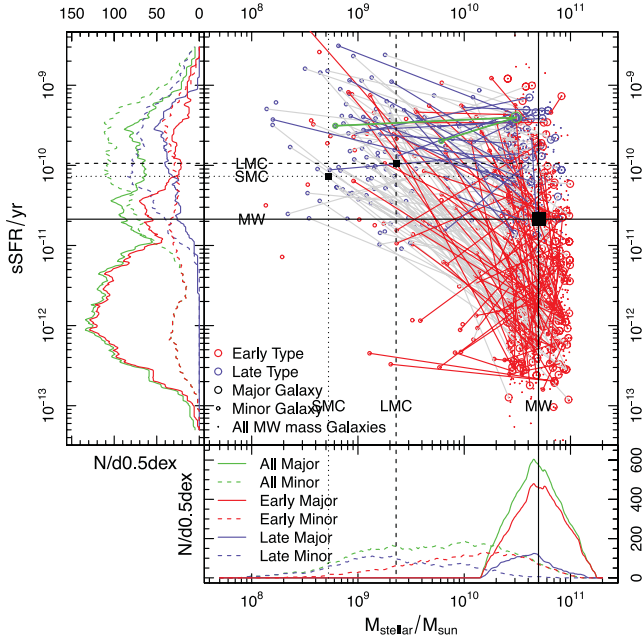
### 6.3.4 SFRs in MW–LMC pairs

Fig. 20 shows the stellar mass versus sSFR of all pairs where the dominant galaxy is within 0.3 dex of the stellar mass of the MW. The links between all visible pairs are shown, where blue lines indicate the galaxies are both late-type, red lines connect early-types and

grey lines are pairs where the type differs. In the redshift range plotted we only expect to be complete down to the LMC, so we will concentrate the discussion here.

Qualitatively we can see that the MW and the LMC are in a well populated part of stellar mass–sSFR parameter space if we just compare to the global populations (i.e. ignore the morphological type of the galaxies). In particular, the LMC is very typical for pair systems in our sample. There are even plenty of examples where the minor galaxy in the pair is more massive than the LMC, yet both the major and minor galaxies are classified as late-type. There is no parameter space where late-type–late-type pair combinations dominate; however, early-type–early-type pairs clearly dominate for low sSFRs.

It is notable that the MW is at the very low extreme for sSFR given it is late-type compared to similar mass dominant pair galaxies. This tallies with our observation in Fig. 11 that the MW has an exceptionally low sSFR given the mass of its halo. Again, the argument that can be made is that the LMC looks broadly typical as a late-type minor pair galaxy (when LMC stellar mass minor pair galaxies exist), but the MW appear to be anomalous as a late-type galaxy given the combination of its stellar mass, halo mass and sSFR. This finding is in agreement with the recent work presented in Robotham et al. (2012), which looked in detail at the occurrence rates of massive satellites around MW stellar mass galaxies. This work also highlighted that the morphological type of the MW and LMC is very unusual – given their stellar masses and the fact they are in a close pair, it is more likely to observe such stellar systems with early-type galaxies.



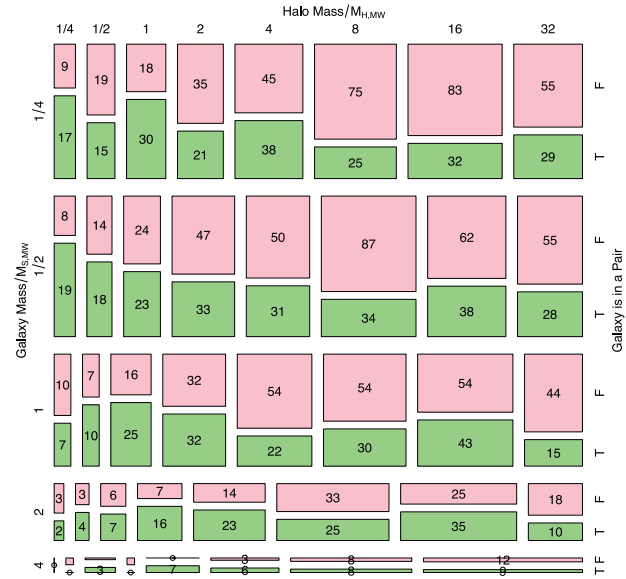
**Figure 20.** Stellar mass versus sSFR of all pairs where the dominant galaxy is within 0.3 dex of the stellar mass of the MW. Also plotted in the background as points are all galaxies that make the mass selection, whether or not they are in pairs. The colouring refers to morphological type (blue = late type; red = early type), and size specifies whether the galaxy is the major (large circle) of minor (small circle) galaxy in the pair, or galaxy that makes the mass selection (point). The coloured lines show connections between galaxy pairs, where blue lines indicate both the major and minor galaxies are late-type, red lines indicate both are early-type, and grey lines show that they differ. Selection shown is the volume sample (see Section 3). Observations: (i) The MW and the LMC are in a well populated part of stellar mass–sSFR parameter space for all galaxies in the sample. (ii) Given it is a late-type dominant pair galaxy, the MW has an unusually low sSFR.

### 6.3.5 Effect of halo mass on close pair rates

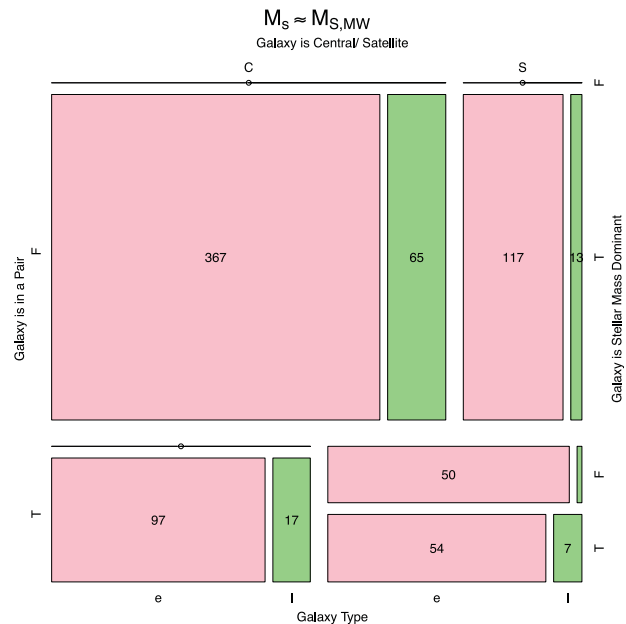
Above we have discussed a number of effects that are stellar mass independent. The dominant effects are caused by pair interactions, with some effects due to the group scale environment also detectable. However, pairs and groups are clearly not mutually independent parts of galaxy parameters space. Indeed it is almost certain that objects detected as pairs will form some part of a galaxy group as found in  $G^3Cv1$  because the compact phase-space selection is usually encompassed by the FoF grouping links used to generate the  $G^3Cv1$ .

Fig. 21 shows the prevalence of galaxy pairs as a function of halo mass and stellar mass. A clear characteristic seen for all stellar masses is that the highest fraction of pairs (by our definition) do not occur in the most massive haloes. In these environments velocity dispersions are large enough that galaxies are unlikely to be observed in the window of dynamical parameter space that we are sensitive to. Instead, the pair rate is highest in lower mass haloes. For galaxies with  $\mathcal{M}_{S,MW} \geq 1$  we see that MW mass haloes host the highest pair fraction of all, i.e. the most common dynamical environment for MW stellar mass galaxies are MW halo mass groups. This lends credence to the idea that the MW and its local halo are a cosmologically significant site for Galactic archeology studies – it is clearly a key environment for understanding the typical experience of stellar mass in large cosmological volumes of the Universe.

Fig. 22 is the final plot tying this work together, showing how combinations of paired and grouped galaxies can predict the



**Figure 21.** Mosaic plot of stellar mass of galaxy (Section 5.2.3)/halo mass (Section 5.2.4)/galaxy in pair (Section 5.1.5). Selection shown is the volume sample (see Section 3). Observations: (i) Galaxy pairs (by the definition of galaxy pair used in this work) are less likely in the highest mass haloes. (ii) The peak in MW–LMC–SMC like pair systems for  $\mathcal{M}_{S,MW} \geq 1$  galaxies is seen at MW mass haloes. (iii) The indication is that MW stellar mass galaxies in MW halo mass groups are key sites of dynamical activity.



**Figure 22.** Mosaic plot of galaxy in pair (Section 5.1.5)/galaxy is central or satellite (Section 5.1.4)/largest galaxy in pair (Section 5.1.6)/type (Section 5.1.1). Subset used for this figure is  $\mathcal{M}_s = \mathcal{M}_{S,MW}$  (Section 5.2.3). In this plot, central galaxies include those not in groups (since they are the central galaxy of their own halo). Selection shown is the mass sample (Section 3). Observations: (i) The late-type fraction is between 10 and 15 per cent for all but the minor pair galaxy subset. (ii) Central galaxies are always the dominant component of a pair, and the late-type fraction does not vary much. Satellite galaxies show a variety of interaction dominance with an almost 54/46 split between being dominant or minor given our observational limits. (iii) For  $\mathcal{M}_s = \mathcal{M}_{S,MW}$  galaxies, the minor galaxy in a pair suppresses the star-forming fraction more than merely being a satellite galaxy of the same stellar mass in a group.

late-type fraction for  $\mathcal{M}_{S,MW}$  mass galaxies. Since we have found that the morphological type displays the strongest associations with other galaxy parameters, this is used as the final contingency. This plot demonstrates that a significant fraction of  $\mathcal{M}_{S,MW}$  mass galaxies are in closely interacting pairs (down to our detectable limits), and those which are the minor galaxy in a galaxy–galaxy interaction have a hugely suppressed late-type fraction (2 per cent). This is a smaller fraction than satellite galaxies not in pairs (10 per cent), or any central galaxy (15 per cent). Central galaxies are always the dominant galaxy in any interaction, so they do not show much residual effect due to pairwise interactions. While central galaxies might always be dominant in pairwise interactions, satellite galaxies show a mixture of being dominant or minor for the selected stellar mass range. Broadly speaking  $\mathcal{M}_{S,MW}$  mass galaxies tend to be the dominant galaxies in all interactions, even when they are in a halo which has a more massive galaxy. In conclusion, the integrated interaction history for satellite galaxies is non-trivial, while central galaxies should possess a much simpler evolution since they will have been much less affected by galaxy–galaxy interactions.

## 7 DISCUSSION AND FUTURE WORK

This work has opened up a number of brackets on the topic of what influences, and is influenced by,  $L^*$  galaxy evolution. Not all of these brackets are closed in this paper, and indeed the expectation is that further analysis and comparisons to detailed simulations will be required to fully untangle the complex causes, effects and mere correlations. There are a few unambiguous remarks that can be made, and these will be discussed first. The more speculative physics hinted at in this work is discussed towards the end of this section.

### 7.1 Clear results

As discussed at length in the introduction, the fate of any single galaxy will be caused by a complex mixture of inputs. In extreme detail the history of any galaxy is certainly unique, and we can only hope to observe broad systematic trends in statistically large samples. The GAMA survey is suitably deep enough and complete enough to separate out the three major influences on galaxy evolution: stellar mass, galaxy–galaxy interactions and halo mass effects.

There are two particularly clear results presented in this work: galaxy stellar mass is a very good predictor of most galaxy properties, and galaxy properties are generally poor predictors of each other.

Finding that stellar mass is the best predictor of simple galaxy observables is very much in line with recent extragalactic literature (e.g. Baldry et al. 2006; Peng et al. 2010; Thomas et al. 2010), where the burgeoning consensus is that stellar mass has the primary correlation with the properties of all galaxies. While we concur with these findings, it is important to stress that the process of assembling stellar mass to form the galaxies is undoubtedly influenced by the environment (e.g. halo) that the galaxy resides in. The explanation can be formed in the language of sub-halo abundance matching (SHAM: e.g. Simha et al. 2012), the halo occupation distribution (HOD: e.g. Brown et al. 2008; Zehavi et al. 2011, 2012) or the conditional luminosity function (CLF: e.g. Cooray 2006), but all share a common theme that the dark matter mass of the parent halo predicts the stellar mass function within the halo. Once the halo has assembled this stellar mass it has a more minor role in dictating the future evolution of the galaxy, and it is certainly true that two

galaxies of a given stellar mass are more likely to share observational properties than two galaxies of a given parent halo mass. However, it is less clear that stellar mass *drives* the evolution. Turning the argument around, it is just as reasonable to state that stellar mass is the end product (or by product) of a large (but bounded) number of evolutionary paths.

While stellar mass appears to be the best predictor of galaxy properties, other galaxy properties are notably poor predictors. While it has long been observed that galaxies possess strong bimodality in a number of properties (see Blanton et al. 2003; Baldry et al. 2004, 2006; Driver et al. 2006; Taylor et al., in preparation), this work has demonstrated quite comprehensively that these bimodalities operate quite independently, with little correlation with each other. The simplest explanation is discussed in the introduction, and posits that differing observational time-scales for a common trigger will naturally allow for galaxy bimodality in many simultaneous dimensions, but with few common properties among individual galaxies. Accurate SED based ‘age’ estimates are not available for GAMA galaxies, so future work to ascertain the role temporal lag has to play in disconnecting various bimodalities will have to be based on unambiguous transition populations, such as E+A galaxies.

Fig. 6 is particularly important in highlighting the danger of oversimplification of galaxy traits. Here we see that much less than half of all  $L^*$  galaxies can be meaningfully described as being ‘red, discless, early type and passive’ and ‘blue, discy, late type and star forming’. In fact, between these four characteristics only discless late types are found to never exist, which must be the case for any reasonable definition of late-type. Every other combination of the 4D contingency space is populated at least to some extent. This alerted us to a key realization: to map out these 4 properties simultaneously with stellar mass, halo mass and pair fraction would be impossible. There are not enough galaxies to properly sample the parameter space, even when using mosaic plots designed for deconstructing multi-dimensional codependencies in relatively small data sets. We chose to focus on investigating star formation and morphological type in latter sections, with the caveat that this will not fully capture all secondary galaxy properties.

### 7.2 Complex results

The analysis presented here is not designed to offer definitive predictions on all mechanisms of galaxy evolution. That said, while stellar mass has been highlighted as the main characteristic that predicts the properties of a galaxy, this work also uncovered strong evidence for secondary evolutionary drivers that act independently of stellar mass. We even found compelling evidence for galaxy–galaxy conformity between observational properties in galaxy groups and galaxy pairs. These secondary drivers produce much weaker effects than stellar mass, so are easily washed away without careful multi-dimensional analysis of the data, which in this work was executed with the use of mosaic plots. The mosaic approach has been used because the statistical power of the data (broadly speaking, the number of galaxies we have to work with) is at the limit for subdivisions into multi-dimensional space. It also allows us to explore discrete and/or binary parameter space not accessible with principal component analysis. The full parameter space that captures all the physics of even just an  $L^*$  galaxy is extremely large, so here we have chosen the few parameters that we assume to be the most important in influencing galaxy properties.

A common theme of the results in this work is that central and satellite galaxies of a given stellar mass *and* halo mass have distinctly different properties – star formation is relatively suppressed

in satellite galaxies, and the late-type fraction is smaller. What is not certain is the mechanism that causes these differences. It is plausible that halo scale effects, for instance the stripping of gas during the accretion of sub-haloes that host the future satellite galaxies, could cause such results. However, the work has been extended to investigate the role that pairwise interactions might have on the same galaxy properties, and many of the same trends were observed. The dominant pair galaxy of a given stellar mass will be more likely to be late type and forming stars than a minor pair galaxy of the same stellar mass. The strength of these differences is of a similar magnitude to the central/satellite variation already noted, and a large fraction of grouped galaxies are undergoing these close galaxy–galaxy interactions.

Previous works have found that environment does play a secondary evolutionary role, after the primary role that stellar mass plays in determining galaxy properties. Peng et al. (2010) note that environment plays an important secondary role in determining the fate of satellite galaxies. This work broadly supports the same view, since we find that ungrouped galaxies display properties most like central galaxies in groups, i.e. it is less a case of the halo strongly influencing central galaxies, and a clearer case that it suppresses star formation in satellite galaxies.

What this work demonstrates is that the mechanism of this ‘environmental’ suppression is heavily associated with galaxy–galaxy interaction physics. The population fractions for a given stellar mass galaxy are altered more dramatically by virtue of being the major or minor galaxy in a pair than by whether the galaxy is a central or satellite in a group. Surprisingly, morphological type is more strongly affected by close interactions than star formation, suggesting the latter is not necessarily clear-cut, with both triggering and suppression physics playing a role. The full role, and the full range of mechanisms, that the environment has on the evolution of galaxies is not a closed case, but some significant fraction of the role that environment plays (beyond assembling stellar mass into galaxies) is simply assembling different stellar mass galaxies within close proximity of other galaxies with velocities conducive to regular close interactions. This concurs with the final conclusion made in Thomas et al. (2010), who highlighted the likely role environment plays in stimulating phase transitions via galaxy–galaxy interactions.

Since we have observed the strong instantaneous effect of galaxy–galaxy interactions, the argument can be extended that the integral of many such interactions over the lifetime of a galaxy could well explain its major properties (after the role of intrinsic stellar mass is considered, of course). The importance of galaxy–galaxy interactions is also suggested by the fact that morphological fractions are equally (if not more) affected by close passages. Feedback mechanisms that do not involve gravity are less able to radically redistribute the orbits of stars (a prerequisite for a morphological transformation), and such events are also able to disturb gas and trigger AGNs and further feedback (e.g. harassment: Moore et al. 1996). Future work includes using simulations to investigate in detail the integrated effects of galaxy–galaxy interactions.

We looked at galaxy conformity of morphological type and emission class in groups and pairs, and found evidence for it, to differing degrees, in both cases. In galaxy groups there was strong evidence for morphological type conformity: i.e. the morphological type of the central galaxy aided the prediction (and was correlated with) the morphological type of the accompanying satellite galaxies. This finding is very much in line with that of Weinmann et al. (2006), with the caveat that they used colour and sSFR to define ‘type’ whereas we used a purely visual classification. However, we found much less compelling evidence for emission class conformity be-

tween central and satellite galaxies. When analysing galaxy pairs in detail we found strong evidence of conformity. In this case, the effect was that the star formation class of one galaxy within a pair could predict its likelihood in the other pair galaxy. The morphological type conformity was considerably weaker, i.e. the parameters that display the strongest conformity swapped between considering central and satellite galaxies in groups, and closely interacting galaxies.

Under the assumption that star formation has a more instantaneous response to stimulation than morphology, this suggests that the morphological type conformity witnessed in galaxy groups is the result of a longer period integrated effect. Such a mechanism could be multiple minor interactions over time. Such dynamical events would slowly rearrange the orbits of the associated galaxies, and the net result of this will be to introduce morphological conformity. On the other hand, the emission class conformity witnessed in galaxy pairs is a transitory response due to a brief close interaction. Only in a minority of cases are the tidal physics of the interaction significant enough to evoke a morphological response in the dynamical window in which we have selected our pairs.

This work was not intended to investigate galaxy occupation statistics in detail. The main analysis considering the typicality of the MW galaxy was presented in Robotham et al. (2012), where the main conclusion was that only  $\sim 11.9$  per cent of MW stellar mass galaxies should host an LMC mass (or more massive) close companion. This bright satellite fraction rate was in close agreement with Boylan-Kolchin et al. (2011). More recent work in Boylan-Kolchin et al. (2011) and Lovell et al. (2012) has suggested that the largest sub-haloes in the MW appear to be unoccupied (a.k.a. the ‘too big too fail’ problem). Further work investigating the faint end of the galaxy group LF is underway, with the intention of quantifying how typical the MW satellites are beyond the LMC regime explored in this work.

## 8 CONCLUSIONS

In this work we have analysed in detail the typical group scale and pair scale environment of galaxies close to the stellar mass of the MW and a factor of 4 either side of it. We did this by visually classifying a complete sample of galaxies that had stellar mass  $M_s > M_{s,MW}/4$ , and all associated pair galaxies and group galaxies with a  $0.01 < z < 0.089$  redshift selection (allowing us to observe down to at least LMC depth in the  $r < 19.4$  mag GAMA survey). The major findings are summarized below.

(i) The correlation between colour, SF, morphological type and presence of a disc is non-trivial, and one parameter cannot be used to accurately predict another. However, some combinations are highly predictive: red, discless early-type galaxies are almost certain to have no emission line features. General relationships are as expected: bluer galaxies are more likely to be strongly star forming than red galaxies etc. For later mosaic plots the most distinctive of these parameters is chosen to avoid information degeneracy (see Fig. 6).

(ii) Larger stellar mass galaxies are more likely to be grouped (in our volume limited definition of group), and also more likely to be a central galaxy if they are in a group. This result in particular falls in line with HOD work conducted over the last decade. These have found that the chance of a massive galaxy being a satellite is small, and indirectly proportional to the halo mass (e.g. Brown et al. 2008; Zehavi et al. 2011, 2012) (see Fig. 7).



(iii) The emission class fractions vary hugely for galaxies of a given stellar mass and group halo mass depending on whether or not it is the central or satellite galaxy. This suggests that galaxy–galaxy interactions play the most important role in shaping the evolution within a group halo, the main role of halo mass being to gather the galaxies together to allow for the interactions (see Figs 7 and 9).

(iv) Galaxy discs are more likely to be present in ungrouped galaxies for galaxies more massive than  $\mathcal{M}_{S,MW}$ , suggesting they can be easily disrupted by dynamical interactions. This has a stronger effect than any enhancement to the disc accretion rate due to there being more material available for minor mergers in galaxy groups. This effect is also enhanced relative to the late-type to early-type fraction (where late-type galaxies always have discs). This relationship is not observed for unpaired galaxies, indicating that the extra degree of isolation offered by being ungrouped, and the longer period to the last major dynamical interaction, is significant in establishing a visible galaxy disc (see Fig. 8).

(v) Satellite galaxies comprise a larger fraction of stellar mass in larger halo mass groups. This is again in line with the standard HOD analysis seen over the last decade (e.g. Brown et al. 2008; Zehavi et al. 2011, 2012) (see Fig. 9).

(vi) There is strong evidence for morphological type conformity in groups, where the morphological type of the central galaxy predicts the morphological type of the satellite galaxy. We also find weaker evidence for galaxy emission class conformity (see Fig. 10).

(vii) The MW is unusually massive, and has an unusually low sSFR, for a late-type galaxy in a MW mass halo (see Fig. 11).

(viii) The LMC is very close to the MW compared to similar mass satellites relative to central galaxies in MW halo mass groups (see Fig. 12).

(ix) If MW mass galaxies are the major component of a pair experiencing a close passage, the star-forming–passive fraction is close to 1–1, but if they are the minor galaxy in of a pair experiencing a close passage then the passive fraction is significantly larger (see Fig. 13).

(x) Massive galaxies are more likely to be in a pair (given our definition of pair, and selection requirements). (See Fig. 14).

(xi) The late-type fraction drops more rapidly with stellar mass than the star-forming fraction, particularly for minor pair galaxies (see Fig. 14).

(xii) The most massive minor pair galaxies show a statistically significant increase in the disc fraction relative to the unpaired case (see Fig. 15).

(xiii) Dominant pair galaxies only find their emission state greatly altered when the pair ratio is close to 1, otherwise we observe the same galaxies as we would in an unpaired system (see Figs 16 and 18).

(xiv) The minor galaxy in a pair is greatly affected by its companion galaxy, and while all star formation is suppressed relative to equivalent stellar mass unpaired galaxies, it becomes lower still when the mass ratio increases (i.e. the major galaxy becomes more massive and dominant within the pair) (see Fig. 17).

(xv) There is strong evidence for galaxy emission class conformity in galaxy pairs, where star-forming emission (or no emission) in one galaxy in a pair strongly predicts the same state in the other galaxy. We find much weaker evidence for morphological type conformity in pairs (see Fig. 19).

(xvi) Given it is a late-type dominant pair galaxy, the MW is unusual in having such a low sSFR (see Fig. 20).

(xvii) Group halo mass has a clear effect on the prevalence of galaxy–galaxy interactions. These become rapidly suppressed in the

largest haloes, and are more common, for a given amount of stellar mass, in lower mass haloes (see Fig. 21).

(xviii) MW–stellar mass interactions peak for MW halo mass groups, suggesting that local Galactic archeology studies will be probing a very important part parameter space for understanding the fate of  $L^*$  galaxies in general (see Fig. 21).

(xix) The observable properties of MW mass galaxies is more radically altered by being the minor pair galaxy in an interaction than merely being the satellite galaxy in a group halo (see Fig. 22).

## ACKNOWLEDGEMENTS

ASGR acknowledges STFC and SUPA funding that was used to do this work. GAMA is a joint European–Australasian project based around a spectroscopic campaign using the Anglo–Australian Telescope. The GAMA input catalogue is based on data taken from the Sloan Digital Sky Survey and the UKIRT Infrared Deep Sky Survey. Complementary imaging of the GAMA regions is being obtained by a number of independent survey programs including *GALEX* MIS, VST KIDS, VISTA VIKING, WISE, Herschel–ATLAS, GMRT and ASKAP providing UV to radio coverage. GAMA is funded by the STFC (UK), the ARC (Australia), the AAO, and the participating institutions. The GAMA website is <http://www.gama-survey.org/>.

## REFERENCES

- Abadi M. G., Moore B., Bower R. G., 1999, *MNRAS*, 308, 947  
 Adelman-McCarthy J. K. et al., 2008, *ApJS*, 175, 297  
 Alpaslan M. et al., 2012, *MNRAS*, 426, 2832  
 Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Z., Lupton R. H., Nichol R. C., Szalay A. S., 2004, *ApJ*, 600, 681  
 Baldry I. K., Balogh M., Bower R. G., Glazebrook K., Nichol R. C., Bamford S. P., Budavari T., 2006, *MNRAS*, 373, 469  
 Baldry I. K. et al., 2010, *MNRAS*, 404, 86  
 Baldry I. K. et al., 2012, *MNRAS*, 421, 621  
 Baldwin J. A., Phillips M. M., Terlevich R., 1981, *PASP*, 93, 5  
 Balogh M. L., Navarro J. F., Morris S. L., 2000, *ApJ*, 540, 113  
 Barnes J. E., Hernquist L., 1992a, *ARA&A*, 30, 705  
 Barnes J. E., Hernquist L., 1992b, *Nat*, 360, 715  
 Bekki K., Couch W. J., Drinkwater M. J., 2001, *ApJ*, 552, 105  
 Benson A. J., Bower R. G., Frenk C. S., Lacey C. G., Baugh C. M., Cole S., 2003, *ApJ*, 599, 38  
 Bertin G., Lin C. C., Lowe S. A., Thurstans R. P., 1989, *ApJ*, 338, 104  
 Blanton M. R., Roweis S., 2007, *AJ*, 133, 734  
 Blanton M. R. et al., 2003, *ApJ*, 594, 186  
 Bournaud F. et al., 2007, *Sci*, 316, 5828  
 Bower R. G., Benson A. J., Malbon R., Helly J. C., Frenk C. S., Baugh C. M., Cole S., Lacey C. G., 2006, *MNRAS*, 370, 645  
 Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., Lemson G., 2009, *MNRAS*, 398, 1150  
 Boylan-Kolchin M., Springel V., White S. D. M., Jenkins A., 2010, *MNRAS*, 406, 2  
 Boylan-Kolchin M., Bullock J. S., Kaplinghat M., 2011, *MNRAS*, 415, L40  
 Brown M. et al., 2008, *ApJ*, 682, 937  
 Cole S., Lacey C. G., Baugh C. M., Frenk C. S., 2000, *MNRAS*, 319, 168  
 Colless M. et al., 2001, *MNRAS*, 328, 1039  
 Cooray A., 2006, *MNRAS*, 365, 842  
 Cooray A., Milosavljević M., 2005, *MNRAS*, 627, L89  
 Croton D. et al., 2006, *MNRAS*, 365, 11  
 Dekel A., Birnboim Y., 2006, *MNRAS*, 368, 2  
 Driver S. P. et al., 2006, *MNRAS*, 368, 414  
 Driver S. P. et al., 2011, *MNRAS*, 413, 971  
 Driver S. P. et al., 2012, *MNRAS*, 427, 3244

- Flynn C., Holmberg J., Portinari L., Fuchs B., Jahreiß H., 2006, MNRAS, 372, 1149
- Freeman K., Bland-Hawthorn J., 2002, ARA&A, 20, 487
- Grootes M. R. et al., 2012, Proc. IAU Symp. 284, The Spectral Energy Distribution of Galaxies. Cambridge Univ. Press, Cambridge, p. 352
- Gunawardhana M. L. P. et al., 2011, MNRAS, 415, 1647
- Gunn J. E., Gott J. R., 1972, ApJ, 176, 1
- Hammer F., Puech M., Chemin L., Flores H., Lehnert M. D., 2007, AJ, 662, 322
- Haywood M., 2008, MNRAS, 388, 1175
- Hill D. T. et al., 2011, MNRAS, 412, 765
- Hopkins A. et al., 2013, MNRAS, preprint (arXiv:1301.7127)
- James P. A., Ivory C. F., 2011, MNRAS, 411, 495
- Kauffmann G., Li C., Zhang W., Weinmann S., 2013, MNRAS, doi: 10.1093/mnras/stt007
- Kelvin L. S. et al., 2012, MNRAS, 421, 1007
- Keres D., Katz N., Weinberg D. H., Dave R., 2005, MNRAS, 363, 2
- Kormendy J., 1977, ApJ, 218, 333
- Kormendy J., Kennicutt R. C., 2004, ARA&A, 42, 603
- Li Y. S., White S. D. M., 2008, MNRAS, 384, 1459
- Liske J., Lemon D. J., Driver S. P., Cross N. J. G., Couch W. J., 2003, MNRAS, 344, 307
- Lovell M. et al., 2012, MNRAS, 420, 2318
- Meyer D., Zeileis A., Hornik K., 2006, J. Stat. Softw., 17, 1
- Moore B., Katz N., Lake G., Dressler A., Oemler A., 1996, Nat, 379, 613
- Nichols M., Colless J., Colless M., Bland-Hawthorn J., 2011, ApJ, 742, 110
- Peng Y. et al., 2010, ApJ, 721, 193
- Read J. I., Lake G., Agertz O., Debattista V. P., 2008, MNRAS, 389, 1041
- Robitaille T. P., Whitney B. A., 2010, ApJ, 710, L11
- Robotham A., Wallace C., Phillipps S., De Propriis R., 2006, MNRAS, 652, 1077
- Robotham A. et al., 2010, PASA, 27, 76
- Robotham A. S. G. et al., 2011, MNRAS, 416, 2640 (R11)
- Robotham A. S. G. et al., 2012, MNRAS, 424, 1448
- Schechter P., 1976, ApJ, 203, 297
- Sellwood J. A., Wilkinson A., 1993, Rep. Prog. Phys., 56, 173
- Simha V., Weinberg D. H., Davé R., Fardal M., Katz N., Oppenheimer B. D., 2012, MNRAS, 423, 3458
- Taylor E. N. et al., 2011, MNRAS, 418, 1587
- Thomas D., Maraston C., Schawinski K., Sarzi M., Silk J., 2010, MNRAS, 404, 1775
- Tollerud E. J., Boylan-Kolchin M., Barton E. J., Bullock J. S., Trinh C. Q., 2011, ApJ, 738, 102
- Verley S., Corbelli E., Giovanardi C., Hunt L. K., 2009, A&A, 493, 453
- Wang W., White S. D. M., 2012, MNRAS, 424, 2574
- Weinmann S. M., van den Bosch F. C., Yang X., Mo H. J., 2006, MNRAS, 366, 2
- Weisz D. et al., 2011, ApJ, 743, 8
- Wijesinghe D. B. et al., 2012, MNRAS, 423, 3679
- Wilkinson M. et al., 2005, MNRAS, 359, 1306
- Zehavi I. et al., 2005, ApJ, 630, 1
- Zehavi I. et al., 2011, ApJ, 736, 59

## APPENDIX A: CONTINGENCY TABLES AND MOSAIC PLOTS

Data that have a large number of parameters can be confusing to visually represent, and most methods that increase the clarity of the data shown reduce the information presented. In astronomy it is not unusual to have data that has multiple discrete columns. A simple example of a set of discrete criteria for galaxies is:

- (i) Is the galaxy in a group (True/False)?
- (ii) Is the galaxy early-type or late-type (Early/Late)?
- (iii) Does the galaxy have a disc (True/False)?
- (iv) What kind of H $\alpha$  emission is observed in the spectra (AGN (A)/star formation (S)/no emission (X))?

An incomplete, but popular, approach to explaining how these different discrete classes depend on each other is to simplify the problem to asking assumed ‘key’ questions such as ‘what fraction of galaxies in groups are late-type?’ The reader will run into difficulty if they happen to ask an unexpected question such as: ‘given a galaxy does not exhibit star formation and has a disc, what is the likelihood that it is early-type and in a group?’ Such a question is perfectly valid, but not precisely communicated by independent bar charts, tables and a few choice examples.

In statistics, parameters that have multiple discrete outcomes are known as contingencies, and hence the relationships between multiple parameters with discrete outcomes are known as contingency tables. A simple hypothetical two-way contingency table is given as an example in Table A1, where the contingencies compared are morphological type (early or late) and presence of a disc (True or False). Table A1 shows the number of objects that exist in each of the four possible logic overlaps. As expected, a large fraction of early-type galaxies do not possess discs, and 100 per cent of late-type galaxies do.

In some scenarios a contingency table is an appropriate form of conveying information. However, to determine what fraction of galaxies are early-type requires the reader to add numbers together that are not given in the Table. In this very simple two-way contingency example this calculation is trivial  $[(310 + 334)/(310 + 334 + 100) = 87$  per cent]; however, simply adding one more contingency makes this task a serious burden.

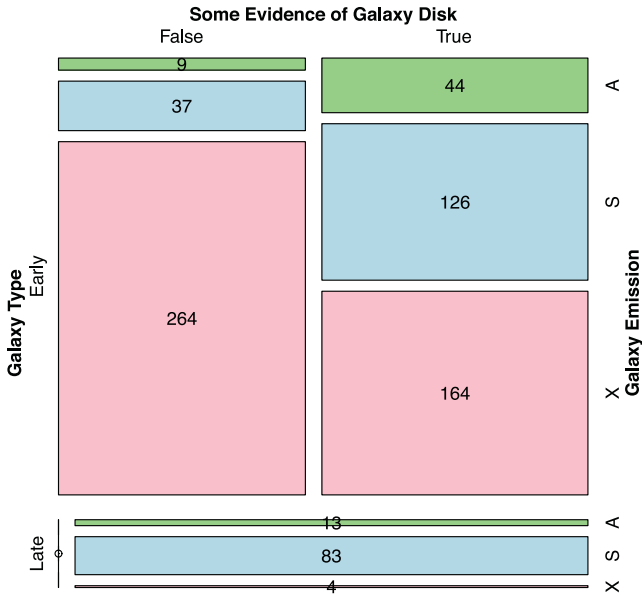
Table A2 expands Table A1 to include an extra contingency (presence of emission line), which has three allowed outcomes: AGN emission (A)/star formation (S)/no emission (X). Calculating the early-type fraction is now quite onerous. It is also harder to get a feel for the data – what is the best predictor of a galaxy possessing strong emission: being early/late type or the presence of a disc? Without a large amount of arithmetic the answer is opaque.

**Table A1.** Two-way contingency table comparing hypothetical numbers of galaxies that are early/late type and do or do not have a disc.

Type	Disc present	
	False	True
Early	310	334
Late	0	100

**Table A2.** Three-way contingency table comparing hypothetical numbers of galaxies that are early/late-type, do or do not have a disc and exhibit star formation (AGN emission (A)/star formation (S)/no emission (X)).

Type	Emission	Disc present	
		False	True
Early	A	9	44
	S	37	126
	X	264	164
Late	A	0	13
	S	0	83
	X	0	4

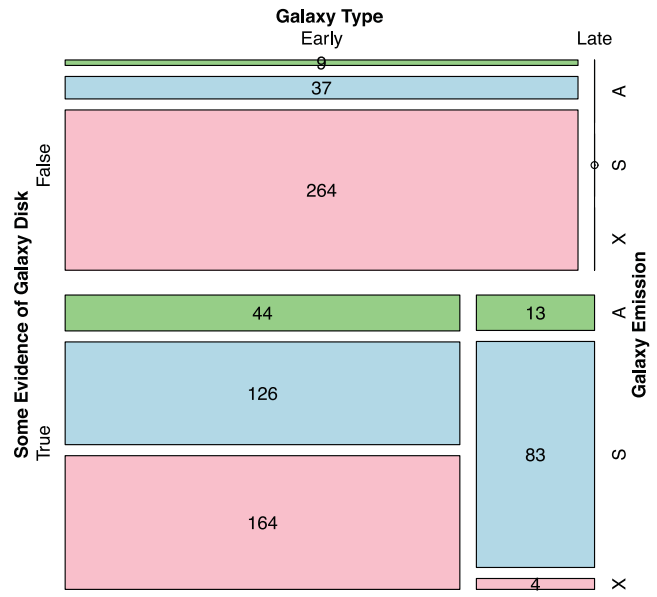


**Figure A1.** Mosaic plot of the data contained in Table A2. The contingency ordering is: type–disc–emission, where cell colouring is according to the emission contingency.

For analysing this sort of multi-dimensional data, with numerous interesting dependencies, statisticians have long favoured mosaic plots. Astronomy has been slow to catch on to this variety of data visualization, but it provides a very powerful means of presenting the data when data sets are large and the author does not want to oversimplify the information.

A mosaic plot works by iteratively sub-dividing the data in a contingency table into ever small cells, alternating whether the splits are made vertically or horizontally. In this example, the first division could be horizontal, showing whether the galaxy is early type or late type (87 per cent to 13 per cent, as calculated above) creating two cells of the appropriate area. The next could then be a vertical split, showing whether the galaxy has any residual disc, so the two cells become four. The last split would again be horizontal, showing the emission class (the final contingency). Since this last contingency has three possible classes, 12 cells would be plotted in total (assuming all cells are occupied).

Fig. A1 shows exactly the plot construction described above, where the counts given in Table A2 are shown within each cell. The combination of cells containing no data (late-type galaxies with no disc) are shown as a line with a small circle drawn on top. The final contingency used to split the rectangles is also used to colour code the data. All cells of the same colour have the same property for the final contingency – in this case all red cells show galaxies with no emission. In such a plot a lot of previously buried information becomes immediately, intuitively, available to the reader. In the subset containing all early-type galaxies we can see an almost identical fraction of galaxies do or do not have discs (without needing to add up any numbers). Also, an early-type galaxy with a disc is more likely to be undergoing strong star formation than an early-type galaxy without a disc. The key to being able to make this last claim is that the relative fraction of galaxies with and without discs is no longer relevant, instead we wish to know the fraction of star-forming galaxies within each sub-class. The mosaic plot allows us to assess the predictive power of possessing a disc by comparing the relative *lengths* of the adjacent rectangles.



**Figure A2.** Mosaic plot of the data contained in Table A2. The contingency ordering is: disc–type–emission, where cell colouring is according to the emission contingency. This is a reversal of the first two contingencies shown in Fig. A1, allowing some information to be extracted more easily, but others harder.

Care must be taken to construct a mosaic plot in the most useful order, since each subsequent division becomes harder to analyse. Fig. A2 shows exactly the same information as Fig. A1, but now the ordering of the first two contingencies is reversed. The questions posed above are no longer as easy to answer with a cursory look at the mosaic since the key cells are no longer side-by-side. However, other observations become easier instead: star-forming galaxies are more common in a late-type galaxy with a disc than an early-type galaxy with a disc. While the ordering is important in terms of what can be easily extracted from the data, the mosaic plot gives access to all the contingency information so all the dependencies can be extracted.

Colour coding according to the final contingency is particularly useful for this reason: a visual summation of the three colours clearly shows the red (i.e. no emission) to be dominant, followed by star forming and lastly AGNs. For gaining rapid access to the most important relationships in multi-dimensional data, mosaic plots are a powerful tool. In terms of information retention, they behave no worse for continuously distributed data than using a running median or other quantiles, this being popular method of displaying relationships in astronomical data with large amounts of intrinsic scatter.

In theory, any number of dimensions of contingency data can be plotted on one mosaic plot- the data is just continually subdivided as outlined above. In practice, it becomes extremely difficult to extract useful information out of more than four dimensions, and the plot labelling becomes confusing (there are four sides to the superset rectangle, so using more than four dimensions in a contingency tables means labelling has to be stacked). For this work, we will restrict the mosaic plots to showing four or fewer dimensions. Also, only one version of each mosaic will be made (i.e. one contingency ordering). This will be chosen as the most instructive of the different combinations attempted.

For a detailed description of the generation of mosaic plots the interested reader should refer to Meyer, Zeileis & Hornik (2006). This

paper discusses the popular ‘vcd’ package for R that was used in this work. This software proved to be the most flexible and extendable of a number tried, and is a good choice for anyone wishing to produce mosaic plots for other data sets. In all the mosaic plots shown in this work the ordering of the mosaic splitting (and thus labelling) is: first contingency, horizontal split, labelling on the left; second contingency, vertical split, labelling on top; third contingency, horizontal

split, labelling on right; fourth contingency, vertical split, labelling on bottom. In cases of crowded labelling, some labels may be omitted/abbreviated/truncated. Cells, or combinations of cells, that sum to zero are represented as a line with a circle drawn on top. Cells with only one count do not have a number printed inside.

This paper has been typeset from a  $\text{\TeX/L\AA\TeX}$  file prepared by the author.