

Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey[★]

Chris J. Lintott,^{1†} Kevin Schawinski,^{1†} Anže Slosar,^{1,2} Kate Land,¹ Steven Bamford,³ Daniel Thomas,³ M. Jordan Raddick,⁴ Robert C. Nichol,³ Alex Szalay,⁴ Dan Andreescu,⁵ Phil Murray⁶ and Jan Vandenberg⁴

¹*Oxford Astrophysics, Denys Wilkinson Building, Keble Road, Oxford OX1 3RH*

²*Berkeley Center for Cosmological Physics, Lawrence Berkeley National Laboratory and Physics Department, Berkeley, CA 94720, USA*

³*Institute of Cosmology and Gravitation, University of Portsmouth, Mercantile House, Hampshire Terrace, Portsmouth PO1 2EG*

⁴*Department of Physics and Astronomy, Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA*

⁵*LinkLab, 4506 Graystone Avenue, Bronx, NY 10471, USA*

⁶*Fingerprint Digital Media, 9 Victoria Close, Newtownards, Co. Down BT23 7GY*

Accepted 2008 July 4. Received 2008 June 26; in original form 2008 April 16

ABSTRACT

In order to understand the formation and subsequent evolution of galaxies one must first distinguish between the two main morphological classes of massive systems: spirals and early-type systems. This paper introduces a project, Galaxy Zoo, which provides visual morphological classifications for nearly one million galaxies, extracted from the Sloan Digital Sky Survey (SDSS). This achievement was made possible by inviting the general public to visually inspect and classify these galaxies via the internet. The project has obtained more than 4×10^7 individual classifications made by $\sim 10^5$ participants. We discuss the motivation and strategy for this project, and detail how the classifications were performed and processed. We find that Galaxy Zoo results are consistent with those for subsets of SDSS galaxies classified by professional astronomers, thus demonstrating that our data provide a robust morphological catalogue. Obtaining morphologies by direct visual inspection avoids introducing biases associated with proxies for morphology such as colour, concentration or structural parameters. In addition, this catalogue can be used to directly compare SDSS morphologies with older data sets. The colour–magnitude diagrams for each morphological class are shown, and we illustrate how these distributions differ from those inferred using colour alone as a proxy for morphology.

Key words: methods: data analysis – galaxies: elliptical and lenticular, cD – galaxies: general – galaxies: spiral.

1 INTRODUCTION

Dividing galaxies into categories based on their morphology, or shapes, has been standard practice since it was first systematically applied by Hubble (1936). It is perhaps surprising that sorting galaxies into categories which are suggested solely by their morphology produces classifications which broadly correlate with other, physical parameters such as the star formation rate or gas fraction. The fundamental distinction drawn is between galaxies with spiral arms

and early-type systems.¹ For most of the 20th century, catalogues of classified galaxies were compiled by individuals or small teams of astronomers (e.g. Sandage 1961; de Vaucouleurs 1991). With the advent of modern surveys [such as the Sloan Digital Sky Survey (SDSS), see Section 1.1] containing many hundreds of thousands of galaxies this approach was no longer practical.

Anticipating the problem these surveys would cause, Lahav et al. (1995) compared classifications from a set of experts who considered a sample of just over 800 galaxies. Their motivation was to create a training set for neural networks, with the aim of automating

[★]This publication has been made possible by the participation of more than 100 000 volunteers in the Galaxy Zoo project. Their contributions are individually acknowledged at <http://www.galaxyzoo.org/Volunteers.aspx>
†E-mail: cjl@astro.ox.ac.uk (CJL); kevins@astro.ox.ac.uk (KS)

¹For the purposes of this paper, we use the term ‘elliptical’ rather than ‘early-type’ as this is the description used on the Galaxy Zoo site. However, the term should be understood as including both elliptical and lenticular systems.

the classification process. While such methods have indeed been developed (Ball et al. 2004), modern studies (e.g. Bernardi et al. 2003; Lintott, Ferreras & Lahav 2006) still separate early-type galaxies from spirals in large data sets by using proxies for morphology rather than by directly determining morphology itself. Typically, selection criteria based on galaxy properties such as colour, concentration index, spectral features, surface brightness profile, structural parameters or some combination of these are used (e.g. Strateva et al. 2001; Abraham, van den Bergh & Nair 2003; Kauffmann et al. 2004; Conselice 2006; Scarlata et al. 2007). Classifications based on structural properties (such as concentration) rather than parameters such as colour which are often but not always correlated are in some sense measures of morphology, and comparison of classifications produced by such methods with smaller data sets is encouraging (Abraham et al. 1996). However, the use of each of these criteria results in an unknown and potentially unquantifiable bias in the resulting sample of galaxies. In other words, although morphological labels are often used for the resulting catalogues, each of these criteria produces a sample different from that obtained by true morphological selection. Comparing results from samples selected using different morphological proxies can therefore be misleading.

Avoiding such confusion between categories is inherently desirable, but is of particular importance – to give just one example – for studies which seek to understand the influence of star formation on the larger scale process of galaxy formation. The colour of a galaxy is often used as a proxy for morphology, but is also a direct consequence of and therefore depends on the star-formation history of the galaxy being studied. By directly classifying objects according to their morphology, the catalogue is sorted according to their dynamics; spirals are rotating, whereas ellipticals are tri-axial (Binney 1982). Other information such as colours, the presence or absence of emission lines, or the galaxy profiles can then be used to investigate the properties of the classified galaxies, rather than being used in the classification itself.

Several subsets of the SDSS have been classified by professional astronomers. Fukugita et al. (2007) recently compiled a catalogue of early-type objects by the visual inspection of ~ 2500 galaxies in the SDSS by three expert classifiers. This is an order of magnitude smaller than the sample used by Schawinski et al. (2007) for their study of active galactic nucleus feedback in early-type galaxies. Their sample (MOSES: MORphologically Selected Ellipticals in SDSS), was obtained by carrying out manual inspection of all objects in the SDSS Data Release 4 (DR4) spectroscopic sample with redshift $0.05 < z < 0.10$ and r -band magnitude $r < 16.8$. The resulting sample consists of 48 023 galaxies, or approximately 5 per cent of the complete SDSS galaxy sample (see below). This sample was then inspected to identify galaxies with an elliptical morphology. The importance of such a morphology-driven classification can be seen from the comparison of MOSES ellipticals with those selected by Bernardi et al (2003). Of the ellipticals selected by Bernardi et al., 5 per cent show emission lines indicative of star-forming activity compared to 18 per cent of the MOSES sample. The sample selected by morphology alone includes a set of star-forming galaxies that are excluded from samples selected by other methods.

Despite the desirability of pure morphological classification, the samples provided by SDSS and other modern surveys are simply too large for astronomers to visually inspect the entire catalogue. Furthermore, without multiple independent classifications of the same galaxy, it is difficult to establish how much confidence can be placed in the classifier or classifiers. Ideally, large numbers of

independent classifications would be made for each galaxy in the sample, allowing the errors to be quantified.

In this paper, we present the results of an attempt to solve this problem by inviting large numbers of people to classify galaxies over the internet. Mass online participation in science was pioneered by projects which made use of idle computers provided by users, the best known of which is SETI@HOME (Werthimer et al. 2001). We aim to make use of the knowledge and skills of our volunteers, rather than just their computers. This solution – known as ‘crowdsourcing’ or ‘citizen science’ – had been successfully employed by projects such as Stardust@Home (Westphal et al. 2006; Mendez 2008). This project was a search for interstellar dust particles in the sample collected by the *Stardust* spacecraft from Comet Wild-2, with the initial selection of samples for further analysis being made by visual inspection. Galaxy Zoo involves an order of magnitude more participants than its predecessors, and is the first attempt to apply these techniques to astrophysical problems. Visual inspection is also an excellent method for serendipitous discovery of the unusual in any data set, and the more unusual objects discovered by Galaxy Zoo classifiers will be discussed in a series of future papers.

1.1 The SDSS

The galaxies for this project were drawn from the SDSS (York et al. 2000). The SDSS is a survey of a large part of the northern sky providing photometry in five filters; u , g , r , i and z (Fukugita 1996), covering approximately 26 per cent of the entire sky. We use the latest available data, contained in Data Release 6 (DR6; Adelman-McCarthy et al. 2008). The SDSS spectroscopic target selection algorithm (Strauss et al. 2002) produces the Main Galaxy Sample, which includes all extended objects with Petrosian magnitude $r < 17.77$ (Petrosian 1976). All objects in this sample which the SDSS photometric pipeline (Lupton et al. 2004) identified as a galaxy were included in the Galaxy Zoo data base, regardless of whether or not such spectra have been obtained to date. This list included a total of 738 175 galaxies drawn from the SDSS main galaxy catalogue. In addition, objects which were not in the spectroscopic catalogue but which had already been observed and as a result classified as a galaxy by the SDSS spectroscopic pipeline were added to our list. This secondary selection comprised 155 037 objects drawn from both the main and luminous red galaxy SDSS catalogues. In all, 893 212 objects were included in our sample. It is reasonable to assume that the accuracy of classification of a galaxy will depend on factors which include the apparent size of the system and its surface brightness. However, as the biases were unquantified before the study was completed, no cuts on this inclusive sample were imposed.

2 GALAXY ZOO

The data for this project was collected via a website.² In order to minimize the degree of knowledge needed by the volunteers, users of the site were not required to distinguish between elliptical and lenticular galaxies or between different classes of spirals (Sa, Sb etc). Visitors to the site were asked to read a brief tutorial giving examples of each class of galaxy, and then to correctly identify a set of ‘standard’ galaxies. These standard systems were selected from the SDSS and classified by team members; those with a low degree of agreement were rejected. Those who correctly classified more than 11 of the 15 standards were allowed to proceed to the main

²<http://www.galaxyzoo.org>



Figure 1. Front page (top) and main analysis page (bottom) from the Galaxy Zoo website.

part of the site. The bar to entry was kept deliberately low in order to attract as many classifiers to the site as possible.

The front page of the site and the main classification page are shown in Fig. 1. SDSS images are shown to volunteers using the ImgCutout web service (Nieto-Santesteban, Szalay & Gray 2004) on the SDSS website (Szalay et al. 2002). The service displays a JPEG cutout image of an area of sky, centred on a galaxy randomly chosen from the sample data base, with an image scale of $0.024R_p$ arcsec per pixel, where R_p is the Petrosian radius for the galaxy. This scaling was empirically chosen so that the entire object was visible, and it appeared as large as possible. While a zoom function may have been useful, the choice of a universal scaling ensures that we are comparing classifications made on a similar basis when considering multiple classifications of the same object by independent classifiers.

These images are colour composites of the three middle filters available in SDSS (g , r and i). Details of the conversion to colour images are given in Lupton et al. (2004). Traditional morphological classifications have used single-band images in order to avoid confusion between morphology and colour. That said, these colour images are particularly suitable for visual classification. In particular, they possess the large dynamic range necessary for the identification of faint features, and have a unique mapping between physical and display colours. The effect of this choice on the data is discussed in Section 4.1.

In addition to sorting galaxies according to their morphology, the website asked classifiers to further divide galaxies they identified as spiral into three subcategories according to the direction of their

Table 1. Galaxy Zoo classification categories showing schematic symbols as used on the site.

Class	Button	Description
1		Elliptical galaxy
2		Clockwise/Z-wise spiral galaxy
3		Anticlockwise/S-wise spiral galaxy
4		Spiral galaxy other (e.g. edge-on)
5		Star or don't know (e.g. artefact)
6		Merger

spiral arms (clockwise/anticlockwise/edge-on). This is a reliable indicator of the sense of the rotation of the galaxy (Pasha & Smirnov 1982). The motivation for this part of the study is twofold. First, we aim to investigate the evidence for a preferred handedness of spiral galaxies reported in the SDSS by Longo (2007). This result conflicts with an earlier paper by Sugai & Iye (1995), who did not find such a result in a different (but comparably sized) data set. Longo's work was based on a sample of 2817 spirals selected by eye from galaxies in the SDSS with a redshift less than 0.04 and a magnitude of $g < 17$. We have been able to extend his analysis to a sample which contains a factor of 10 more galaxies, and the results are presented in a companion paper (Land et al. 2008). Secondly, it will also prove possible to use our results to calculate the two point correlation function for rotating spirals, an interesting new constraint on models of galaxy formation (Slosar et al., in preparation).

Each object extracted from the SDSS data base was thus classified as belonging to one of six categories: spiral (clockwise rotation), spiral (anticlockwise rotation), spiral (edge-on/rotation unclear), elliptical, merger or star/don't know. The symbols used for this classification are shown in Table 1. In order to keep the task as simple as possible, no further distinction was made between barred and unbarred spiral systems, for example. Once a classification is chosen, then the image of the next galaxy is automatically displayed.

As the Galaxy Zoo website gathers data, these are stored into a live STRUCTURED QUERY LANGUAGE (SQL) data base. For each entry, we store the timestamp, user identification, galaxy identification and the classification chosen by the user. Classifications by unregistered visitors are discarded and the user requested to register and complete the tutorial described above. For the analysis, this data base may be downloaded and processed through the pipeline described below.

Although some classifiers will inevitably know (or will learn from experience during the project) that spiral galaxies tend to be bluer than elliptical galaxies, the tutorial stressed that objects should be classified according to their morphology alone. No mention was made of the colour–morphology relation. In order to quantify the effect of colour on our results, a selection of monochrome images was introduced to the sample, and the results are discussed in Section 4.1. The data discussed in this paper represent the final results from the first stage of the project; Galaxy Zoo 2, which will ask for more detailed classifications, will follow shortly.

3 PRODUCING A CATALOGUE

The website was successful in attracting large numbers of classifiers and classifications, as shown in Figs 2 and 3. Each galaxy in our sample was thus viewed and classified multiple times, with a mean of ~ 38 classifications per galaxy. A variety of strategies are available

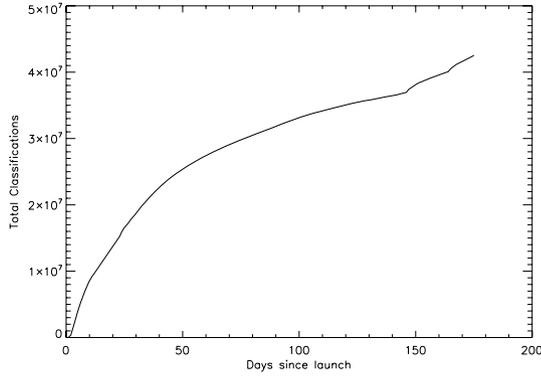


Figure 2. Cumulative classifications collected by the Galaxy Zoo site. The sudden increases visible at ~ 145 and ~ 160 d correspond to email newsletters being sent out to those registered with the site. These led to a sustained increase in the rate of classification. Following day 140, data collected contributed to the bias study described in Section 4.1.

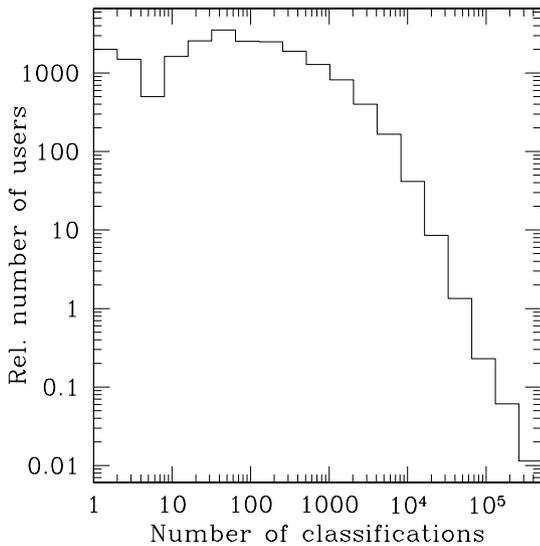


Figure 3. The distribution of classifications among users. A small number have completed more than 100 000 classifications each, while the peak of the distribution is ~ 30 classifications per user.

to convert from these raw classifications to a final catalogue. In this section we compare the results from several different strategies.

The first step in data reduction involves removing obviously bogus classifications. A small number of users seem to have recorded a number of these classifications, either using some sort of automated mechanism or due to some unknown problem with their browser. They are easy to discern by the fact that they have multiple classifications for a small number of galaxies. We find all users which have classified two or more galaxies more than five times each. This is extremely unlikely by Poisson distribution and hence *all* data points from such users are discarded. There are 36 such potentially malicious users, amounting to less than 0.05 per cent of the total number of participants. Furthermore, in order to account for accidental double clicks, if a user has classified the same galaxy more than once, we take into account only the first classification from each user. This latter stage ensures that no single user can unduly influence the classification assigned to a single galaxy. The two steps of this cleaning process together remove about 4 per cent of our data set.

In the next step, we create the so called *combined spirals* sample. In this sample, we combine all three possible spiral classifications into a single classification. This is useful for studies that require just a simple split into elliptical and spiral samples. All the subsequent analysis is performed on both the separated spirals and combined spirals samples.

We are then in a position to create the unweighted final sample. The simplest method involves giving all classifiers equal weight and simply calculating the distribution of classification for each galaxy. This distribution can now be interpreted in a Bayesian manner: it represents our state of knowledge about that particular galaxy.

Using these classifications, we test the effect of seeing and sky brightness on our classifications, and find no significant dependence over the limited variation over which they vary in the SDSS photometry.

3.1 Weighted sampling

The unweighted method discussed above does not discriminate between results from those who think carefully before classifying each galaxy, and those who take less time. Neither does it distinguish between the ability of our classifiers. It may therefore make sense to attempt to identify particularly ‘good’ users. The meaning of ‘good’ is naturally subjective, but one obvious strategy is to pay more attention to classifications from users who tend to agree with the majority. For this analysis, each user of the website was initially assigned a unit weight, as in the unweighted sample described above. A preliminary classification could then be obtained as before for each galaxy. The weighting assigned to individual users could then be adjusted according to how they agree with this assessment. A new set of galaxy classifications could then be prepared using the new weights, and the process repeated until the weights converge.

In order to avoid the weightings assigned to users being distorted by the fainter end of our galaxy sample, we used only galaxies with Petrosian radius $r_p > 4.5$ arcsec and $r < 17$ for our weighting. This leaves 257 000 galaxies involved in producing user weightings, although the resulting weightings are applied to all galaxy classifications.

The algorithm used was as follows: let the weight of a user, k , be w_k and set all initial weights to one. We then integrate the data base to find $h_i(j)$, the number of users who classified galaxy i as being class j (elliptical, anticlockwise spiral etc. . .). $N_g(k)$ is the number of galaxies classified by user k . The weights and h_i are then updated by using the formulæ

$$h_i(j) = A \sum_{(k = \text{users who voted } j \text{ for galaxy } i)} w_k, \quad (1)$$

where A is chosen so that the mean user weight is one, and

$$w_k = \sum_i \frac{h_i(j \text{ chosen by user } k \text{ for galaxy } i)}{N_g(k)}. \quad (2)$$

This process can then be repeated until convergence. The final product is the weighted sample of galaxy classifications and a set of user weights.

It should be noted, however, that the process of reweighting favours the majority opinion. A user that is most similar to other users will get upweighted and an user that does not conform to the pattern will get downweighted. However, the overall agreement between users does not necessarily mean improvement as people can agree on a wrong classification. These effects must be calibrated using comparison with standardized observations as described in Section 4.

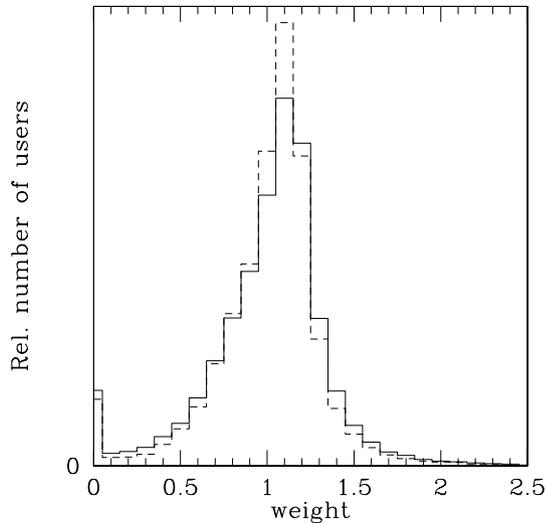


Figure 4. The distribution of user weights for the separated (solid line) and combined (dashed line) data sets. The distribution for the separated spirals is slightly wider than that for the separated spirals sample.

The distribution of user weights for both the separated and combined spiral samples is shown in Fig. 4. Both distributions are slightly skewed towards the low-weighted end, and the combined spiral distribution is tighter than that for the separated spiral data. This reflects the fact that as there are fewer possibilities to choose from in classifying a galaxy, for a set number of classifications better signal to noise is obtained, allowing us to better constrain the user weights.

There are thus four possible combinations of separated spirals/combined spirals and weighted/unweighted samples. Unless otherwise stated, we use the weighted sample. For each sample, we distil the data further into clean and superclean samples. The galaxy is in a clean or superclean sample if it has more than 10 votes (in practice, this applies to almost our entire sample) and if 80 or 95 per cent of users (or user weights in the case of the weighted sample), respectively, agree on its type. These are extremely strong limits; an 80 per cent agreement is the equivalent of a 5σ detection for a galaxy with 10 classifications, or a 10σ detection for a galaxy with the mean number of classifications.

Examples of objects in each category randomly extracted from the weighted superclean sample are shown in Fig. 5, and examples from the clean sample in Fig. 6. It should be noted that the combined spiral sets cannot be recovered simply by taking the single spirals clean set and combining classes 2, 3 and 4. For example, a galaxy that has all its votes evenly split between classes 2 and 3 (clockwise and anticlockwise spirals) will definitely be included in the combined spiral clean set, but would not appear in the separated spirals clean set.

The effect of this weighting process is shown in Table 2 for the separate spirals, and in Table 3 for the combined spirals. These tables show that in all cases the vast majority (above 99 per cent in all but two cases) of classifications made in the unweighted samples are carried forward into the weighted sample. This means that the weighting described above is not changing classifications. However, extra galaxies are included in each classification in the weighted sample. This effect is largest (~ 15 per cent) for the elliptical classes; this can easily be explained by the fact that it is more difficult to agree on the presence of spiral structure than on an elliptical morphology. A smaller proportion of spiral systems reach the stringent criteria

for inclusion in the superclean sample. Both clean and superclean samples have a surprisingly small fraction of mergers; re-analysis by a member of the team has suggested that the 80 per cent weighted vote criteria required for inclusion in the clean sample is too low for the mergers category. The majority of those systems with more than 60 per cent weighted vote in this category appear to be mergers, as do a substantial number of systems with between 20 and 60 per cent weighted vote. This reflects the reluctance of our volunteers to click ‘merger’, which is possibly due to the lack of prominence given to the button (see Fig. 1). For consistency, we continue to use the 80 per cent threshold in this paper, and defer further discussion to future work (Darg et al., in preparation). In total, more than 300 000 galaxies are included in the most inclusive sample; this is the largest sample of morphologically classified galaxies by a factor of 10.

Inspection of Tables 2 and 3 immediately reveals that many more galaxies in the clean sample have been classified as elliptical than spiral. The elliptical–spiral ratio is ~ 3 for both the weighted and unweighted clean sample. The combined spiral clean sample produces a much lower ratio (~ 2). This difference is another illustration of the discrimination against spirals discussed in the previous paragraph. The combined spirals data should be free of such effects, but still has a large elliptical fraction. This reflects the tendency of our users to classify objects which are faint, small in angular extent on the sky or both as elliptical if no spiral features are visible. Unless they are particularly ‘discy’ such galaxies will be classified as elliptical regardless of their true nature. This effect means that it is therefore important to apply magnitude cuts to the data before using data for the population as a whole; individual users of the Galaxy Zoo data will require different cuts and so we do not impose any on the clean sample ourselves. The elliptical fraction for a volume-limited sample of well-classified galaxies is discussed in Section 5.

4 COMPARISON WITH OTHER SAMPLES

In order to assess the reliability of the Galaxy Zoo classifications, we compare our sample with that produced by previous projects. The MOSES sample (Schawinski et al. 2007) described in Section 1 consists of 15 729 galaxies classified as elliptical selected from an initial set of 48 023 galaxies. Of the 48 023 the clean sample includes classifications for 19 649 systems. The results for the weighted clean sample are given in Table 4.

More than 99.9 per cent of the galaxies classified as MOSES ellipticals which are in the Galaxy Zoo clean sample are found to be ellipticals by Galaxy Zoo. However, ~ 15 per cent of the ellipticals included in both the Galaxy Zoo clean sample and MOSES were not classified as elliptical by MOSES. All MOSES ellipticals in the superclean sample are classified by Galaxy Zoo as ellipticals, but the sample contains 3 per cent more ellipticals than MOSES. These extra ellipticals are the result of the different motivation of the studies; MOSES was an attempt to produce a very clean set of ellipticals, whereas the Galaxy Zoo samples include more of the S0–Sa continuum in the resulting sample. The Galaxy Zoo instructions to volunteers did not mention discs at all, and so galaxies which are elliptical in morphology but have visible discs would have been included in Galaxy Zoo but not in MOSES.

There are also ellipticals which are in the MOSES catalogue but not in the clean sample. The distribution of weighted votes for MOSES ellipticals including both those included in the clean sample and those which are not is shown in Fig. 7. The majority of weighted votes in almost all cases support an elliptical classification. The requirement for the clean classification of a weighted vote of 80 per cent thus lies in the middle of a continuous distribution

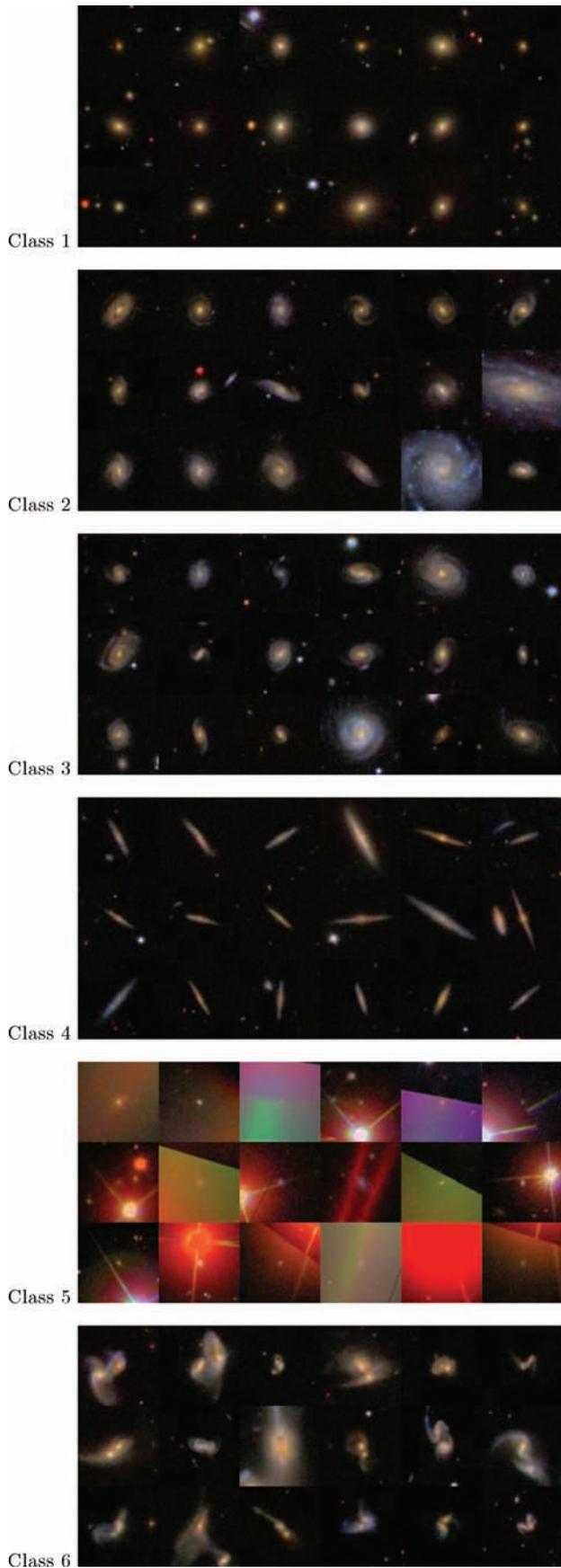


Figure 5. Examples of galaxies in each class drawn from the weighted superclean sample. Each image is 51.2×51.2 arcsec².

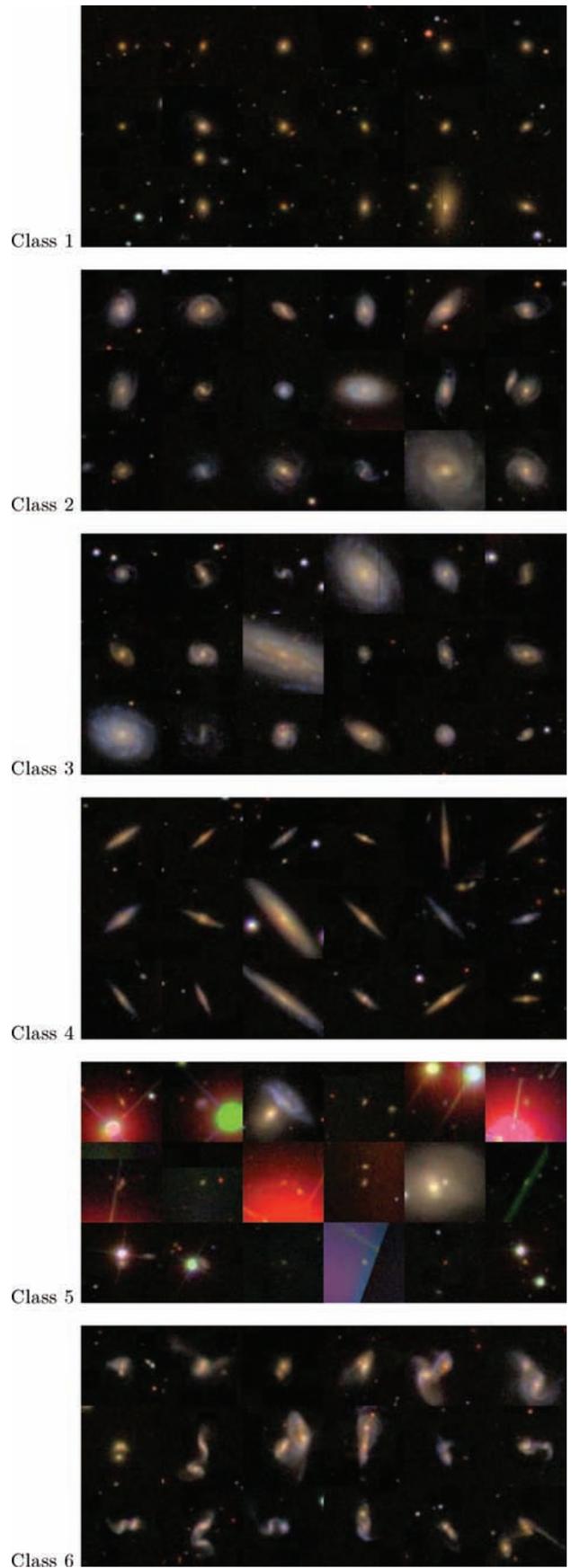


Figure 6. Examples of galaxies in each class drawn from the weighted clean sample. Each image is 51.2×51.2 arcsec².

Table 2. Comparison of classification between weighted and unweighted samples. For each class, the table shows the number of galaxies so classified, the percentage increase in weighted over unweighted classifications and the percentage of the unweighted sample in common with the weighted sample.

Sample	Class	No. in weighted	No. in unweighted	Per cent increase	Per cent common
Clean	Elliptical	219 326	184 743	15.8	99.98
Clean	CW spiral	17 571	17 100	2.7	99.70
Clean	ACW spiral	18 946	18 471	2.5	99.72
Clean	Other spiral	27 310	26 037	4.7	99.46
Clean	Star/don't know	8134	8074	0.7	99.75
Clean	Merger	1062	961	9.5	99.48
Superclean	Elliptical	26 200	19 121	37.0	99.7
Superclean	CW spiral	6532	6106	7.0	99.4
Superclean	ACW spiral	7486	7034	6.4	99.4
Superclean	Other spiral	4760	4247	12.1	99.2
Superclean	Star/don't know	5589	5393	3.6	99.5
Superclean	Merger	70	62	12.9	96.8

Table 3. As Table 2, but for the combined spirals data set.

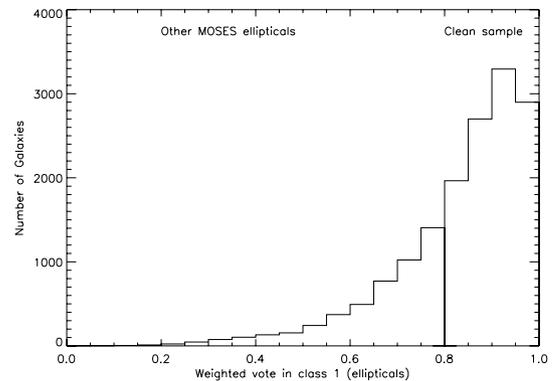
Combined spirals sample	Class	No. in weighted	No. in unweighted	Per cent increase	Per cent common
Clean	Elliptical	208 437	184 743	12.8	99.9
Clean	Spiral	101 855	97 848	4.1	99.9
Clean	Star/don't know	8126	8074	0.6	99.9
Clean	Merger	1056	961	9.9	99.5
Superclean	Elliptical	23 806	19 121	24.5	99.7
Superclean	Spiral	34 673	32 559	6.5	99.7
Superclean	Star/don't know	5573	5393	3.3	99.7
Superclean	Merger	67	62	8.1	96.8

Table 4. Comparison of classifications for galaxies in both MOSES and the Galaxy Zoo weighted clean sample. Most MOSES ellipticals are classified by Galaxy Zoo as elliptical.

	Moses e	Moses other	Moses all
Elliptical	10 858	1676	12 534
ACW spiral	0	2493	2493
CW spiral	2	2,598	2600
Other spiral	4	1940	1944
Star/don't know	0	4	4
Merger	4	70	74
Total	10 868	8781	19 649

of weights. In most cases, the remaining votes show that a small minority of users selected other options, usually for good reasons such as the presence of a nearby satellite trail or some evidence of a disturbed morphology. The weighted vote in the spiral categories is below 20 per cent in all but an insignificant number of cases. This example thus illustrates the stringency of the clean and superclean samples; only galaxies on which a large majority of users agree are included in the final samples.

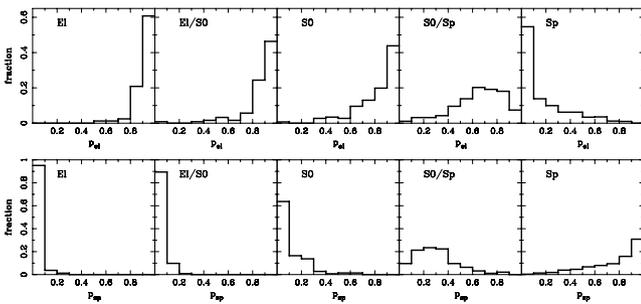
In order to investigate this effect and provide an independent check on the data, we consider another set of SDSS galaxies, those classified by Fukugita et al. (2007). They use the statistic ‘T’ as their classification, taken from an average of three classifiers (rounding to the nearest half integer). The options available are: 0(E), 1(S0), 2(Sa), 3(Sb), 4(Sc), 5(Sd) and 6(Im). Unlike MOSES, therefore, we can use this smaller sample to probe the response

**Figure 7.** Weighted vote in class 1, corresponding to ellipticals, for the 15 729 galaxies classified in the MOSES sample as elliptical. Those with a weight in this class greater than 80 per cent are included in the Galaxy Zoo clean sample, but it is clear from this figure that this is an effectively arbitrary cut-off point. For approximately 90 per cent of the galaxies, the majority of weighted votes are in the elliptical category.

of Galaxy Zoo users to galactic morphologies of a wide variety of subtypes. $1 \leq T \leq 5$ are ‘spiral’ systems, $T = 0, 0.5$ are ‘elliptical’ and $T < 0$ and $T = 6$ are unclassified or irregular systems. Of their sample of 2275 galaxies, we have clean classifications for 1300 galaxies (621 are included in the superclean sample). The mean T for galaxies included in the clean sample and classified as elliptical is 0.52, and that for spirals 2.54. A full comparison for

Table 5. Comparison of the combined spirals clean (top) and superclean (bottom) sample results with those from Fukugita et al. (2007). Their classification is given on the x -axis, and the Galaxy Zoo results on the y -axis. See Table 1 for details of our classification system.

$T =$	<0	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	all
Elliptical	1	267	190	170	41	11	2	0	0	0	0	0	0	0	682
Spiral	3	0	0	0	5	21	71	136	151	160	38	13	5	2	605
Star/don't know	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1
Merger	5	0	1	0	0	1	1	2	0	1	1	0	0	0	12
Total	9	268	191	170	46	33	74	138	151	161	39	13	5	2	1300
$T =$	<0	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	all
Elliptical	0	145	88	52	3	0	0	0	0	0	0	0	0	0	288
Spiral	0	0	0	0	1	7	22	58	94	119	24	6	2	0	333
Star/don't know	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Merger	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total	0	145	88	52	4	7	22	58	94	119	24	6	2	0	621

**Figure 8.** Histograms showing a comparison between Galaxy Zoo classifications and those from Fukugita et al. The axis plots the fraction of the weighted vote in the clean sample that Galaxy Zoo allocated to elliptical (top) and spiral (bottom) for those galaxies classified by Fukugita et al. as elliptical (EI), elliptical/S0 (EI/S0), S0, S0/spiral (S0/Sp) and spiral (Sp).

the clean sample is given in Table 5, and the distribution of weights shown in Fig. 8.

The vast majority of galaxies classified as elliptical in the clean sample are classified as elliptical ($T = 0, 0.5$) by Fukugita et al. Of the ellipticals in the clean sample 92 per cent correspond to early-type galaxies in the Fukugita et al. sample (S0 or ellipticals). The equivalent figure for the superclean sample is 99 per cent. All but two of the remaining galaxies classified by Galaxy Zoo as ellipticals were classified by Fukugita et al. as Sa. This supports the hypothesis that the excess of ellipticals seen when comparing to the MOSES sample is composed mostly of Sa galaxies; astronomers are more reluctant than the general public to classify something with a definite disc as an elliptical galaxy. No mention of discs was made in the instructions to our classifiers, but such an addition is an obvious change to make in future versions of the website. For the ellipticals, we find no obvious trend between T and magnitude, but T appears to be correlated with the weight of the classification. Switching to the superclean sample therefore improves the correlation between the samples.

Finally, Longo (2007) selected the spiral galaxies used in his study by visual inspection of galaxies in the SDSS. Of the 2834 galaxies in this sample, 2498 are included in the Galaxy Zoo clean sample, 2491 of which are classified as spirals. The other seven are classified as mergers (4) or as ‘star/don’t know’ (3). A comparison with the clean separate spiral catalogue finds excellent agreement for the winding sense of the spiral arms (99.6 per cent). In the 10 cases where there was a disagreement, further inspection reveals that

the disagreement can in each case be put down to human error in the catalogue of Longo (2007), illustrating the advantage of obtaining multiple independent classifications for each system.

The three data sets with which we have compared Galaxy Zoo were compiled in very different ways to test different hypotheses. However, in each case we find a remarkable degree of agreement (better than 90 per cent in most cases) between our data set and those compiled by professional astronomers. We can therefore conclude that using data from volunteers will not substantially degrade the quality of the resulting catalogue while expanding the number of classified galaxies by a large factor.

4.1 Measuring bias

The aim of the Galaxy Zoo study was to produce a catalogue of morphologically selected galaxies, independent of the bias introduced by using proxies for morphology. Potentially, the strongest of these biases is the correlation between morphology and colour. While the instructions to users did not include any mention of colour, it is a fact that most spirals are significantly bluer than most elliptical galaxies, and this fact will be quickly learnt by classifiers. It is therefore possible that our selection will include a residual colour bias. In order to quantify the size of any such effect, a programme of bias testing was undertaken. Users were shown either a mirror image of the original data or a monochrome image produced from the coloured images. (These images are not single filter images, but rather a black and white version of the colour image produced by the SDSS pipeline as described above.) The results are given in Table 6.

Any bias study such as this runs the risk of changing the behaviour of those taking part itself, a phenomenon known in social science as the Hawthorne effect (Mayo 1933; Adair, Sharpe & Huynh 1989). To give just one example of how this might affect Galaxy Zoo,

Table 6. Results of the bias study. The numbers given are the average percentage of votes that each class receives per galaxy, with 1σ errors obtained from jackknife resampling (see Land et al. for details).

Class	Original (per cent) (σ)	Monochrome (per cent) (σ)	Mirrored (per cent) (σ)
1	53.82 (0.12)	55.96 (0.12)	55.02 (0.12)
2 & 3 & 4	32.37 (0.13)	28.97 (0.13)	30.05 (0.13)
5	10.12 (0.06)	11.06 (0.06)	11.26 (0.06)
6	3.69 (0.05)	4.01 (0.05)	3.67 (0.05)

users may be more cautious with their classifications if they think that they are being tested for bias rather than just being asked to make their best guess.

A change in user behaviour between the original classifications and those collected as part of this bias study is indeed seen. In particular, users are more careful in their classifications during the bias study. This effect makes it impossible to make a fair comparison between classifications made before the bias study started and those collected during it. However, we do not expect mirroring the images to influence the choice between spiral and elliptical galaxies, and we can thus use the mirrored images as a control. The result of a comparison between classifications of monochrome and mirrored images is a significant (of order 5σ) difference in behaviour. Users shown monochrome images are more likely to classify a galaxy as an elliptical, and correspondingly less likely to classify a galaxy as a spiral. There is also a bias in favour of classifying a galaxy as a merger; this is presumably due to the loss of colour information which enables us to distinguish two separate galaxies from one merging system. However, although these are statistically significant differences, they are small. The mean percentage of votes for the elliptical class increases from 55 to 56 per cent, for example. We are thus justified in ignoring this bias when using the catalogues for most purposes.

By using the monochrome images as a control, we can test for a bias in the classification of the direction of spiral arms. A significant bias in favour of anticlockwise classifications was found, and is discussed in Land et al. (2008). This bias is important even for clean and superclean samples. For example, in the clean sample Land et al. report (clockwise, anticlockwise) numbers of (839, 932) in monochrome images and (864, 905) for mirrored images. The errors are found to be ± 25 galaxies via the jackknife sampling method, and the reported excess of anticlockwise galaxies is consistent with the level of bias measured. If a galaxy is required to be in the clean sample in both monochrome and mirrored samples then a sample of 739 clockwise and 739 anticlockwise is obtained.

We also expect a bias towards elliptical galaxies for more distant systems as it becomes harder to resolve features which would indicate a spiral system. Providing a conservative cut in magnitude, size or redshift (or some combination of the three) is made, then this bias can be safely ignored. When considering the properties of the population as a whole, it is possible to be more quantitative in accounting for the effect of this bias on the results; for a full discussion of this technique, see Bamford et al. (2008).

5 COLOUR-MAGNITUDE DIAGRAMS

In Figs 9 and 10, we show the colour-magnitude diagrams for those galaxies in our superclean sample which have spectroscopic magnitudes. The magnitudes and colours are based on absolute magnitudes calculated using `KCORRECT v4_1_4` (Blanton & Roweis 2007). The elliptical galaxies in the sample have a mean $u - r$ of 2.55, significantly redder than the spirals (mean $u - r = 1.85$). These results correspond to the classic ‘red sequence’ of early-type galaxies found by previous studies (e.g. Sandage & Visvanathan 1978; Bower, Kodama & Terlevich 1998), with the blue galaxies existing not on a tightly defined sequence but rather in a ‘blue cloud’. The division between the two is not straightforward, however. For example, close inspection of Fig. 10 reveals that the sample contains populations of both blue elliptical galaxies (which are discussed in a companion paper to this, Schawinski et al. 2008) and red spirals (the morphology–density relation for which is shown in Bamford et al. 2008 and which will be discussed in a future paper).

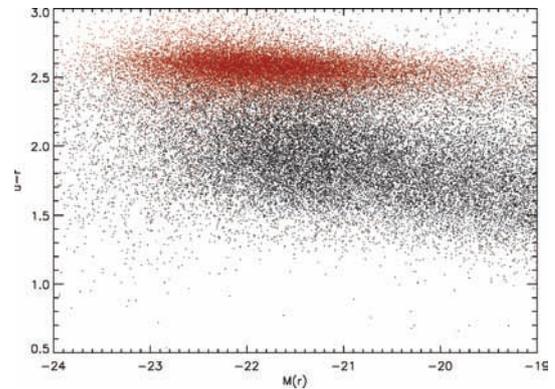


Figure 9. Colour–magnitude diagram for galaxies in the weighted superclean combined spirals sample. Systems classified as spiral are shown in black, those classified as elliptical in red.

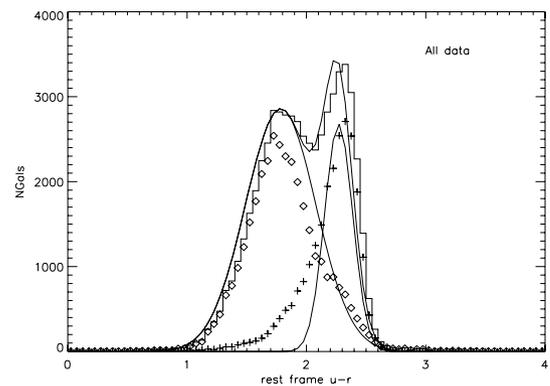


Figure 10. Colour distributions for galaxies in the clean combined spirals sample. Crosses mark ellipticals, diamonds spirals. A two-Gaussian fit to the complete data is shown (top), together with the individual Gaussians used in the fit. The curve shown is the limit for the main galaxy catalogue; objects below this line were drawn from the luminous red galaxy (LRG) sample.

In particular, Fig. 10 includes a fit to the data with two Gaussians. The rest frame colours used in these plots are calculated using k -corrections from Blanton & Roweis (2007). The combined result is reasonable, but as expected from the discussion above the two Gaussians do not clearly divide spiral from elliptical galaxies. In particular, the ‘blue elliptical’ population forms a substantial contribution to the blue side of the redder of the two Gaussians. This result illustrates the importance of true morphological classification; even a sophisticated division between ‘red’ and ‘blue’ systems will not entirely separate the two morphological types.

In order to explore further the properties of our sample in colour-magnitude space, we construct three volume-limited subsamples from those objects in the clean sample for which spectroscopic redshifts have been obtained. In order to improve confidence in the data, samples were constructed both for $r < 17.77$ (solid lines) and $r < 17.0$ (dashed lines). The cuts applied are illustrated in Fig. 11. The most luminous sample is dominated by elliptical galaxies, with a elliptical–spiral ratio of 1.99. The intermediate sample has a ratio of 0.98, and is thus evenly split between the two classes, whereas the sample including the faintest galaxies is dominated by spirals, with a ratio of 0.57.

Colour–magnitude diagrams for each of these subsamples are shown in Fig. 12. We also show Gaussian fits to the data based on

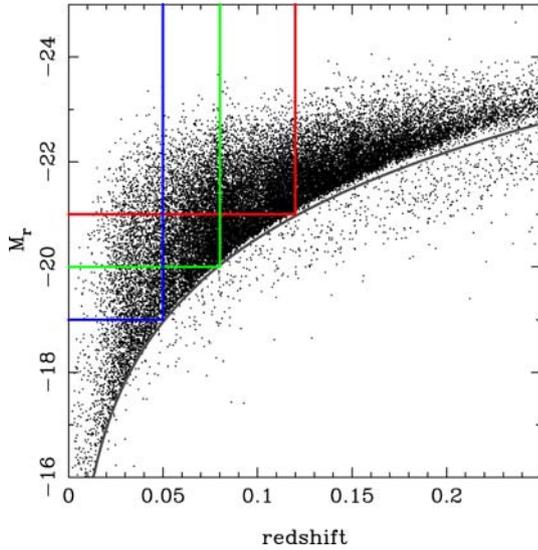


Figure 11. Cuts applied to create volume limited subsamples from the clean sample. The curve shows the $r = 17.77$ line converted to M_r , using the distance modulus but neglecting k -corrections, and corresponds to the main galaxy sample limit for an object with a flat spectrum. Points below (and some just above) this line are drawn from the LRG sample.

those in Baldry et al. (2004). Baldry et al. divide galaxies drawn from the SDSS into red and blue systems, defining a galaxy as red if $C'_{ur} > C_{ur}$, where C_{ur} is the rest-frame (k -corrected) $u - r$ colour and

$$C'_{ur} = 2.06 - 0.244 \tanh\left(\frac{M_r + 20.07}{1.09}\right). \quad (3)$$

Fits to the data shown in Fig. 12 are Gaussians with the same mean and variance as those derived in Baldry et al. These Gaussians were then normalized to fit our data set.

We show the results in Fig. 12, where some general trends are immediately apparent. The proportion of galaxies classified as ellipticals is larger in the sample which includes only the most luminous galaxies. The results also confirm as before that in none of the three cases can the two distributions (red and blue galaxies) which would be derived in the absence of morphological information be simply interpreted as corresponding to ‘early’ and ‘late’ type galaxies. It is not possible to define a single colour with which to divide the two classes of galaxy; rather the distributions overlap to a large extent.

The biggest difference between the populations inferred from Gaussian fitting and those obtained by our morphological classification is the presence of a substantial number of red galaxies which were classified as spiral systems in the lower luminosity samples. In fact, the population of galaxies which we classify as morphologically spiral contains a substantial number of systems with $u - r$ colours greater than ~ 2.2 . Many of these systems may actually be lenticulars; however, distinguishing a well-resolved edge-on S0 galaxy from an edge-on spiral is impossible by visual classification alone. Despite this contamination, however, true red spirals do exist in the data and morphological and colour bimodality are – at least for this intriguing population – decoupled. There is a corresponding population of blue elliptical galaxies, the most extreme examples of which are the subject of a companion paper (Schawinski et al. 2008), but they are less significant here.

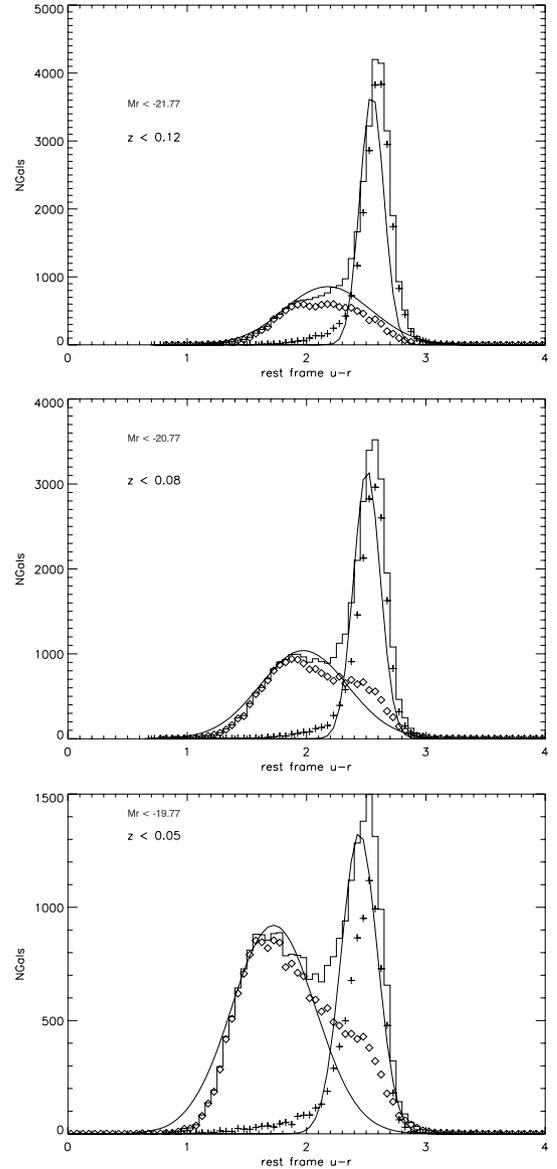


Figure 12. Colour distributions for our set of three volume limited samples. Petrosian magnitudes are used, and k -corrections and absolute magnitudes derived from spectroscopic redshifts. As in Fig. 10, crosses represent ellipticals, diamonds spirals and the histogram the combined data set. Gaussian fits were made to the combined data (spirals and ellipticals) but only the individual Gaussians are shown.

6 CONCLUSION

We have described Galaxy Zoo, a web-based project which invited the public to classify galaxies imaged as part of the SDSS. By combining the classifications of more than 100 000 participants in the largest astronomical collaboration in history, we are able to produce catalogues of simple galaxy morphology which agree with those compiled by professional astronomers to an accuracy of better than 10 per cent. Our results thus suggest that the general public can reliably classify large sets of galaxies with a similar accuracy to professional astronomers. The largest of the Galaxy Zoo catalogues includes more than 300 000 galaxies reliably classified at more than 5σ confidence according to morphology, a factor of 10 larger than previous work. Due to the repeated, independent classifications of the same object it is possible to quantify the errors in

the classification, and produce catalogues of differing fidelity for different purposes (such as the clean and superclean catalogues discussed here). By examining a volume-limited subset of the data in colour–magnitude space, we illustrate the differences between the colour and morphological bimodalities in the data. The presence of a substantial number of red galaxies classified as spiral in particular underlines the importance of morphological classification; our results show that a traditional morphological classification cannot be reproduced by cuts on colour alone.

ACKNOWLEDGMENTS

In addition to the contribution from Galaxy Zoo volunteers, we also acknowledge invaluable contributions from Edd Edmondson, Pete Wilton, Alice Sheppard and Danny Locksmith. We thank Professor Joe Silk for his encouragement. CJL acknowledges support from the STFC Science in Society Programme. We would also like to thank our referee, Roberto Abraham, for his constructive comments.

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society, and the Higher Education Funding Council for England. The SDSS Web Site is <http://www.sdss.org/>.

The SDSS is managed by the Astrophysical Research Consortium for the Participating Institutions. The Participating Institutions are the American Museum of Natural History, Astrophysical Institute Potsdam, University of Basel, University of Cambridge, Case Western Reserve University, University of Chicago, Drexel University, Fermilab, the Institute for Advanced Study, the Japan Participation Group, Johns Hopkins University, the Joint Institute for Nuclear Astrophysics, the Kavli Institute for Particle Astrophysics and Cosmology, the Korean Scientist Group, the Chinese Academy of Sciences (LAMOST), Los Alamos National Laboratory, the Max Planck Institute for Astronomy (MPIA), the Max Planck Institute for Astrophysics (MPA), New Mexico State University, Ohio State University, University of Pittsburgh, University of Portsmouth, Princeton University, the United States Naval Observatory and the University of Washington.

REFERENCES

Abraham R. G., van den Bergh S., Glazebrook K., Ellis R. S., Santiago B. X., Surma P., Griffiths R. E., 1996, *ApJS*, 107, 1
 Abraham R. G., van den Bergh S., Nair P., 2003, *ApJ*, 588, 218
 Adair J. G., Sharpe D., Huynh, C.-L., 1989, *Rev. Ed. Res.*, 59, 215
 Adelman-McCarthy J. K. et al., 2008, *ApJS*, 175, 297
 Baldry I. K., Glazebrook K., Brinkmann J., Ivezić Ž., Lupton R. H., Nichol R. C., Szalay A. S., 2004, *ApJ*, 600, 681

Ball N. M., Loveday J., Fukugita M., Nakamura O., Okamura S., Brinkmann J., Brunner R. J., 2004, *MNRAS*, 348, 1038
 Bamford S. P. et al., 2008, *MNRAS*, submitted (arXiv:0805.2612)
 Bernardi M. et al., 2003, *AJ*, 125, 1817
 Binney J., 1982, *ARA&A*, 20, 399
 Blanton M. R., Roweis S., 2007, *AJ*, 133, 734
 Bower R. G., Kodama T., Terlevich A., 1998, *MNRAS*, 299, 1193
 Conselice C. J., 2006, *MNRAS*, 373, 1389
 Fukugita M., Ichikawa T., Gunn J. E., Doi M., Shimasaku K., Schneider D. P., 1996, *AJ*, 111, 1748
 Fukugita M. et al., 2007, *AJ*, 134, 579
 Hubble E., 1936, *The Realm of the Nebulae*. Yale Univ. Press, New Haven
 Kauffmann G., White S. D. M., Heckman T. M., Ménard B., Brinchmann J., Charlot S., Tremonti C., Brinkmann J., 2004, *MNRAS*, 353, 713
 Lahav O. et al., 1995, *Sci*, 267, 859
 Land K. et al., 2008, *MNRAS*, 388, 1686
 Lintott C. J., Ferreras I., Lahav O., 2006, *ApJ*, 648, 826
 Longo M., 2007, arXiv:0707.3793
 Lupton R., Blanton M. R., Fekete G., Hogg D. W., O’Mullane W., Szalay A., Wherry N., 2004, *PASP*, 116, 133
 Mayo E., 1933, *The Human Problems of an Industrial Civilization*, ch. 3. MacMillan, New York
 Mendez B. J. H., 2008, in Garmany C., Gibbs M. G., Moody J. W., eds, *ASP Conf. Ser. Vol. 389, EPO and a Changing World: Creating Linkages and Expanding Partnerships*. Astron. Soc. Pac., San Francisco, p. 219
 Nieto-Santisteban M., Szalay A., Gray J., 2004, in Ochsenbein F., Allen M., Egret D., eds, *ASP Conf. Ser. Vol. 314, Astronomical Data Analysis & Software Systems XIII*. Astron. Soc. Pac., San Francisco, p. 666
 Pasha I. I., Smirnov M. A., 1982, *Ap&SS*, 86, 215
 Petrosian V., 1976, *ApJ*, 209, L1
 Sandage A. R., 1961, *The Hubble Atlas of Galaxies*. Carnegie Institute of Washington, Washington
 Sandage A. R., Visvanathan N., 1978, *ApJ*, 225, 742
 Scarlata C. et al., 2007, *ApJS*, 172, 406
 Schawinski K., Thomas D., Sarzi M., Maraston C., Kaviraj S., Joo S.-J., Yi S. K., Silk J., 2007, *MNRAS*, 382, 1415
 Schawinski K. et al., 2008, *MNRAS*, submitted
 Strateva I. et al., 2001, *AJ*, 122, 1861
 Strauss M. A. et al., 2002, *AJ*, 123, 1810
 Sugai H., Iye M., 1995, *MNRAS*, 276, 327
 Szalay A. S., Gray J., Thakar A. R., Kunszt P. Z., Malik T., Raddick M. J., Stoughton C., vandenBerg J., 2002, in *Proc. 2002 ACM SIGMOD Int. Conf. on Management of Data*, p. 570
 de Vaucouleurs G., de Vaucouleurs A., Corwin H. G., Jr, Buta R. J., Paturel G., Fouque P., 1991, *Third Reference Catalog of Bright Galaxies*. Springer-Verlag, New York
 Werthimer D. et al., 2001, *Proc. SPIE*, 4273, 104
 Westphal A. J. et al., 2006, *AGU, Fall Meeting*, abstract P52B-08
 York D. G. et al., 2000, *AJ*, 120, 1579

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