Mon. Not. R. Astron. Soc. **000**, 1–29 (2013) Printed 29 May 2013 (MN LATEX style file v2.2)

Galaxy Zoo 2: detailed morphological classifications for 304,122 galaxies from the Sloan Digital Sky Survey

Kyle W. Willett[1⋆], Chris J. Lintott[2,3], Steven P. Bamford[4], Karen L. Masters[5,6], Brooke D. Simmons[2], Kevin Schawinski[7], Lucy Fortson[1], Ramin A. Skibba[8], Kevin R.V. Casteels[9], Edward M. Edmondson[5], Arfon M. Smith[2,3], M. Jordan Raddick[9], Sugata Kaviraj[2,10], Thomas Melvin[5], Robert J. Simpson[2], Robert C. Nichol[5,6], William C. Keel[12]

[1] *School of Physics and Astronomy, University of Minnesota, USA*
[2] *Department of Physics, University of Oxford, UK*
[3] *Astronomy Department, Adler Planetarium and Astronomy Museum, USA*
[4] *School of Physics and Astronomy, University of Nottingham, UK*
[5] *Institute of Cosmology and Gravitation, University of Portsmouth, UK*
[6] *SEPnet, South East Physics Network, UK*
[7] *Institute for Astronomy, ETH, Zürich, Switzerland*
[8] *Center for Astrophysics and Space Sciences, University of California San Diego, USA*
[9] *Departament d'Astronomia i Meteorologia, Universitat de Barcelona, Spain*
[10] *Department of Physics and Astronomy, Johns Hopkins University, USA*
[11] *Centre for Astrophysics Research, University of Hertfordshire, UK*
[12] *Department of Physics and Astronomy, University of Alabama, USA*

Accepted XXXXXXX

ABSTRACT

We present the data release for Galaxy Zoo 2 (GZ2), a citizen science project with more than 16 million morphological classifications of 304,122 galaxies drawn from the Sloan Digital Sky Survey. Morphology is a powerful probe for quantifying a galaxy's dynamical history; however, automatic classifications of morphology (either by computer analysis of images or by using other physical parameters as proxies) still have drawbacks when compared to visual inspection. The large number of images in current surveys make visual inspection of each galaxy impractical for individual astronomers. GZ2 uses classifications from volunteer citizen scientists to measure morphologies for all galaxies in the DR7 Legacy survey with $m_r > 17$, in addition to deeper images from SDSS Stripe 82. While the original Galaxy Zoo project only identified galaxies as early-types, late-types, or mergers, GZ2 provides details of finer morphological features. These include the presence of bars, bulges, and the shapes of edge-on disks, as well as quantifying the relative strengths of galactic bulges and spiral arms. This paper presents the full public data release for the project, including measures of accuracy and bias. We show that the majority of GZ2 classifications agree with those made by professional astronomers, especially for T-types, strong bars, and arm curvature. Both the raw and reduced data products can be obtained in electronic format at `http://data.galaxyzoo.org`.

Key words: catalogues, methods: data analysis, galaxies: general, galaxies: spiral, galaxies: elliptical and lenticular

1 INTRODUCTION

The Galaxy Zoo project (Lintott et al. 2008) was launched in 2007 to provide morphological classifications for nearly one

⋆ E-mail: willett@physics.umn.edu

million galaxies drawn from the Sloan Digital Sky Survey (SDSS; York et al. 2000). This scale of effort was made possible by combining classifications from hundreds of thousands of volunteers, but in order to keep the task at a manageable level of complexity only simple morphological distinctions were initially requested, primarily dividing systems into elliptical, spiral and merger. Following the success of the original project, we wanted to determine if the same method could be used for a more complex classification system. This paper presents data and results from Galaxy Zoo's successor, called Galaxy Zoo 2 (GZ2), comprising detailed morphologies for more than 300,000 of the largest and brightest SDSS galaxies.[1]

While the morphological distinction used in the original Galaxy Zoo (GZ1) – that which divides spiral and elliptical systems – is the most fundamental, there is a long history of enhanced classifications. The first systematic approach to classification (Hubble 1936) included a division between barred and unbarred spirals, creating the famous 'tuning fork' diagram. Further distinctions were based on the shape of early-type systems or tightness of late-type spiral arms. These finer distinctions are often believed to be correlated with physical parameters of the systems being studied; the presence of a bar, for example, may drive gas inwards and be correlated with the growth of a central bulge (reviews are given in Kormendy & Kennicutt 2004 and Masters et al. 2011). Similarly, the presence of a central bulge is likely to indicate a history of mass assembly through significant mergers (Martig et al. 2012 and references therein). Careful classifications of morphological features are thus essential if the assembly and evolution history of galaxies is to be fully understood.

Whereas traditional morphological classification relied on the careful inspection of small numbers of images by experts (e.g., Sandage 1961; de Vaucouleurs et al. 1991), the sheer size of modern data sets make this approach impractical. Detailed classifications of SDSS images by experts have been done by both Fukugita et al. (2007) and Baillard et al. (2011), who determined modified Hubble types for samples of 2253 and 4458 galaxies, respectively. The largest detailed professional effort to date was undertaken by Nair & Abraham (2010), who provide classifications of 14,034 galaxies. Galaxy Zoo 2 includes more than an order of magnitude more systems, each with a large number of independent inspections. The size of this sample allows for a more complete study of small-scale morphological features and better statistics for rarer classes of objects, while multiple classifications per object yields an estimate of the associated uncertainty.

The use of proxies for morphology such as colour, concentration index, spectral features, surface brightness profile, structural features, spectral energy distribution or some combination of these is not an adequate substitute. Each proxy has an unknown and possibly biased relation with the morphological features under study. With a sufficiently large set of galaxies, however, we can fully sample the morphological diversity of the local population and quantify the relationship between morphology and the proxies discussed above.

Despite recent advances in automated morphological classification, driven in part by the availability of large training sets from the original Galaxy Zoo (Banerji et al. 2010; Huertas-Company et al. 2011; Davis & Hayes 2013), the state of the art does not provide an adequate substitute for classification by eye. In particular, as Lintott et al. (2011) note, such efforts typically use proxies for morphology as their input, and so they suffer equally from the objections raised above to the use of morphological proxies. The release of the dataset associated with this paper will be of interest to those developing such machine learning and computer vision systems.

These results were made possible by the participation of hundreds of thousands of volunteer 'citizen scientists'. The original Galaxy Zoo demonstrated the utility of this method in producing both large-scale catalogues as well as serendipitous discoveries of individual objects (see Lintott et al. 2011 for a review of Galaxy Zoo 1 results). Since then, this method has been expanded beyond simple shape classifications to include supernova identification (Smith et al. 2011), exoplanet discovery (Fischer et al. 2012; Schwamb et al. 2012) and a census of bubbles associated with star formation in the Milky Way (Simpson et al. 2012).

Several results based on early Galaxy Zoo 2 data have already been published. Masters et al. (2011, 2012) use GZ2 bar classifications to show a clear increase in bar fraction for galaxies with redder colours, lower gas fractions, and more prominent bulges. Hoyle et al. (2011) developed a separate interface to measure bar properties from GZ2-selected objects, showing that the bars themselves are both redder and longer in redder disk galaxies. Skibba et al. (2012) demonstrated that a significant correlation exists between barred and bulge-dominated galaxies at separations from 0.15–3 Mpc. Kaviraj et al. (2012) used GZ2 to study early-type galaxies with visible dust lanes, while Simmons et al. (2013) discovered a population of AGN host galaxies with no bulge, illustrating how black holes can grow and accrete via secular processes. Finally, Casteels et al. (2013) quantify morphological signatures of interaction (including mergers, spiral arms, and bars) for galaxy pairs in the SDSS.

This paper is organised as follows. Section 2 describes the sample selection and method for collecting morphological classifications. Section 3 outlines the data reduction process, and Section 4 describes the tables that comprise the public data release. Section 5 is a detailed comparison of GZ2 to four additional morphological catalogues that were created with SDSS imaging. Section **??** presents morphologically-sorted colour-magnitude diagrams as an example of the science that can be done with GZ2. We summarise our results in Section 6.

This paper uses the WMAP9 cosmological parameters of $H_0 = 71.8$ km/s/Mpc, $\Omega_m = 0.273$, and $\Omega_\Lambda = 0.727$ (Hinshaw et al. 2012).

2 PROJECT DESCRIPTION

2.1 Sample selection

The primary sample of objects used in Galaxy Zoo 2 comprise approximately the brightest 25% of the resolved galaxies in the SDSS North Galactic Cap region. Our sample was restricted to the SDSS Data Release 7 (DR7) 'Legacy' catalogue (Abazajian et al. 2009), and therefore excludes ob-

[1] `http://zoo2.galaxyzoo.org`

servations made by SDSS for other purposes, such as the SEGUE survey. Spectroscopic targets came from the SDSS Main Galaxy Sample (Strauss et al. 2002).

Several cuts on the data were applied to the DR7 Legacy sample for GZ2 selection. The goal was to include only the nearest, brightest, and largest systems for which fine morphological features should be resolved and classifiable. We required a Petrosian half-light magnitude brighter than 17.0 in the r-band (after Galactic extinction correction was applied), along with a `petroR90_r`, the radius containing 90% of the r-band Petrosian aperture flux, greater than 3 arcsec. Galaxies which had a spectroscopic redshift in the DR7 catalogue outside the range $0.0005 < z < 0.25$ were removed; however, galaxies without reported redshifts were kept. Finally, objects which are flagged by the SDSS pipeline as SATURATED, BRIGHT or BLENDED without an accompanying NODEBLEND flag were also excluded. The 245,609 galaxies satisfying all these criteria are referred to as the "original" sample.

An error in the selection query meant that the "original" sample initially missed some objects, specifically those flagged as both BLENDED and CHILD. These galaxies, which are typically slightly brighter, larger and bluer than the general population, were added to the GZ2 site on 2009-09-02. These additional 28,174 galaxies are referred to as the "extra" sample.

In addition to galaxies from the DR7 Legacy, GZ2 also classified images from Stripe 82, a multiply-imaged section along the celestial equator in the Southern Galactic Cap. The selection criteria were the same as for the Legacy galaxies, with the exception of a fainter magnitude limit of $m_r < 17.77$. For the Stripe 82 sample only, GZ2 includes multiple images of individual galaxies: one set of images at single-depth exposures, plus two sets of co-added images from multiple exposures. Coadded images combined 47 (south) or 55 (north) separate scans of the region, resulting in an object detection limit approximately two magnitudes lower than in normal imaging (Annis et al. 2011).

The primary sample for GZ2 analysis consists of the combined "original", "extra", and Stripe 82 normal-depth images with $m_r \leqslant 17.0$. We have verified that there are no significant differences in the classifications between these sub-samples (possibly due to a time-dependent bias, for example) and thus can be reliably used as a single data set. This is hereafter referred to as the GZ2 **main sample** (Table 1), and is used for the bulk of the analysis in this paper. Data from both the Stripe 82 normal-depth images with $m_r > 17.0$ and the two sets of coadded images are separately included.

2.2 Image creation

Images of galaxies from the Legacy and Stripe 82 normal depth surveys were generated from the SDSS ImgCutout web service (Nieto-Santisteban, Szalay & Gray 2004). Each image is a *gri* colour composite 424×424 pixels in size, scaled to ($0.02\times$`petroR90_r`) arcsec per pixel.

Coadded images from Stripe 82 were generated from the corrected SDSS FITS frames. Frames were combined using Montage (Jacob et al. 2010) and converted to a colour image using a slightly modified version of the asinh stretch code (Lupton et al. 2004), with parameters adjusted to replicate the normal SDSS colour balance. The parameterisation of the stretch function used is:

$$f(x) = \mathrm{asinh}(\alpha Q x)/Q \quad (1)$$

where $Q = 3.5$ and $\alpha = 0.06$. The colour scaling is [1.000,1.176,1.818] in g, r and i, respectively.

The first set of coadded images were visually very different from the normal SDSS images. Changing the colour balance to maximise the visibility of faint features, however, resulted in more prominent background sky noise; since each pixel is typically dominated by a single band, the background is often brightly coloured by the Lupton et al. (2004) algorithm. Due to concerns that this noise would be an obvious sign that the images were from deeper data (potentially biasing the classifications), we created a second set of coadd images in which the colour of background pixels was removed. This was achieved by reducing the colour saturation of pixels outside of a soft-edged object mask.

The original and desaturated coadd image sets are labeled "stripe82_coadd_1" and "stripe82_coadd_2", respectively (Table 1). Subsequent analysis revealed very few differences between the classifications for the images using the two coadd methods (see Section 4.2).

Sample	N_{gal}	N_{class} median	m_r [mag]
original	245,609	44	17.0
extra	28,174	41	17.0
Stripe 82 normal	21,522	45	17.77
Stripe 82 normal ($m_r < 17$)	10,188	45	17.0
Stripe 82 coadd 1	30,346	18	17.77
Stripe 82 coadd 2	30,339	21	17.77
main original + extra + S82 ($m_r < 17$)	283,971	44	17.0

Table 1. Basic properties of the galaxy samples in GZ2, including the total number of galaxies (N_{gal}), the median number of classifications per galaxy (N_{class}), and the apparent magnitude limit.

2.3 Decision tree

Data for Galaxy Zoo 2 was collected via a web-based interface. Volunteers needed to register with a username for their classifications to be recorded.

Classification begins with the user being shown an SDSS colour composite image of a galaxy alongside a question and set of possible responses dealing with the galaxy's shape, similar to Galaxy Zoo 1. Further morphological classification then proceeds via a multi-step decision tree. In this paper, we define a *classification* as the total amount of information collected about one galaxy by a single user completing the decision tree. Each individual step in the tree is a *task*, which consists of a *question* and a finite set of possible *responses*. The selection of a particular response is referred to as the user's *vote*.

Classification begins with a slightly modified version of the GZ1 task, identifying whether the galaxy is either "smooth", has "features or a disk", or is a "star or artifact".

The appearance of subsequent tasks depend on the user's previous responses. For example, if the user clicks on the "smooth" button, they are subsequently asked to classify the roundness of the galaxy; this task would not be shown if they had selected either of the other two responses.

The GZ2 tree has 11 classification tasks with a total of 37 possible responses (Figure 1 and Table 2). A classifier selects only one response for each task, after which they are immediately taken to the next task in the tree. Tasks 01 and 06 are the only questions that are always completed for a given object. Once a classification is complete, an image of the next galaxy is automatically displayed and the user can begin classification of a new object.

Data from the classifications were stored in a live Structured Query Language (SQL) database. In addition to the morphology classifications, the database also recorded a timestamp, user identifier, and image identifier for each classification.

2.4 Site history

Galaxy Zoo 2 was launched on 2009-02-16 with the "original" sample of 245,609 images. The "extra" galaxies from the Legacy survey were added on 2009-09-02. The normal-depth and first coadded Stripe 82 images were mostly added on 2009-09-02, with an additional $\sim$ 7700 of the coadded images added on 2010-09-24. Finally, the second version of the coadded images were added to the site on 2009-11-04.

For most of the duration of GZ2, the images shown to classifiers were selected from the database in a random order. We wanted to ensure, however, that each galaxy ultimately had enough classifications to accurately measure its uncertainty. Therefore, in the final period of GZ2, accompanied by a competition with a running tally (called the "Zoonometer"), objects with low numbers of classifications were shown at a higher rate. The "stripe82_coadd_1" sample was removed from the site at this time. The main sample galaxies finished with a median of 44 classifications; the minimum was 16 classifications, and $> 99.9\%$ of the sample had at least 28. The "stripe82_coadd_2" galaxies had a median of 21 classifications and $> 99.9\%$ had at least 10 (Figure 2).

The last GZ2 classifications were collected on 2010-04-29, with the project spanning just over 14 months. The archived site continued to be maintained, but classifications were no longer recorded. The final dataset contained 16,340,298 classifications (comprising a total of 58,719,719 tasks) by 83,943 volunteers.

Task	Question	Responses	Next
01	*Is the galaxy simply smooth and rounded, with no sign of a disk?*	smooth	07
		features or disk	02
		star or artifact	**end**
02	*Could this be a disk viewed edge-on?*	yes	09
		no	03
03	*Is there a sign of a bar feature through the centre of the galaxy?*	yes	04
		no	04
04	*Is there any sign of a spiral arm pattern?*	yes	10
		no	05
05	*How prominent is the central bulge, compared with the rest of the galaxy?*	no bulge	06
		just noticeable	06
		obvious	06
		dominant	06
06	*Is there anything odd?*	yes	08
		no	**end**
07	*How rounded is it?*	completely round	06
		in between	06
		cigar-shaped	06
08	*Is the odd feature a ring, or is the galaxy disturbed or irregular?*	ring	**end**
		lens or arc	**end**
		disturbed	**end**
		irregular	**end**
		other	**end**
		merger	**end**
		dust lane	**end**
09	*Does the galaxy have a bulge at its centre? If so, what shape?*	rounded	06
		boxy	06
		no bulge	06
10	*How tightly wound do the spiral arms appear?*	tight	11
		medium	11
		loose	11
11	*How many spiral arms are there?*	1	05
		2	05
		3	05
		4	05
		more than four	05
		can't tell	05

Table 2. The GZ2 decision tree, comprising 11 tasks and 37 responses. The 'Task' number is an abbreviation only and does *not* necessarily represent the order of the task within the decision tree. The texts in 'Question' and 'Responses' are displayed to volunteers during classification, along with the icons in Figure 1. 'Next' gives the subsequent task for the chosen response.

3 DATA REDUCTION

3.1 Multiple classifications

In a small percentage of cases, individuals classified the same image more than once. In order to treat each vote as an independent measurement, such multiple classifications were removed from the data, keeping only their votes from the last submitted classification. Repeat classifications occurred for only $\sim$ 1% of all galaxies. The removal of the repeats only altered the final vote fractions (thus changing the morphological classification) for $\lesssim 0.01\%$ of the sample.

3.2 Consistency and individual user weighting

The next step is to reduce the influence of potentially unreliable classifiers (whose classifications are consistent with random selection). To do so we applied an iterative weighting scheme, similar to that used for GZ1. First, we calculated the vote fraction ($f_r = n_r/n_t$) for every response to every task for every galaxy, weighting each user's vote equally. Here, n_r is the number of votes for a given response and n_t is the total number of votes for that task. Each vote is compared to the vote fraction to calculate a user's consistency κ:

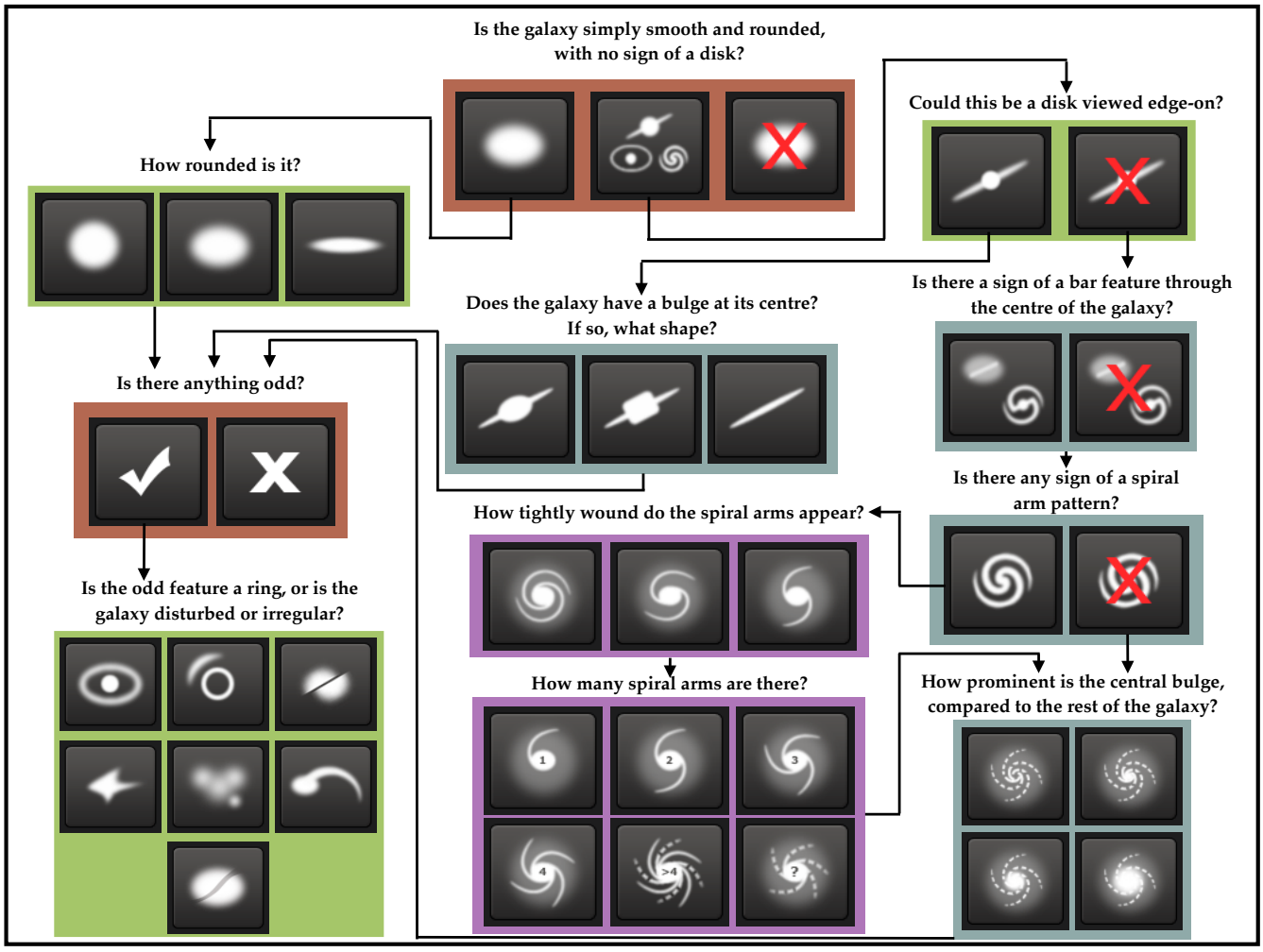

Figure 1. Flowchart of the classification tasks for GZ2, beginning at the top centre. Tasks are colour-coded by their relative depths in the decision tree. Tasks outlined in brown are asked of every galaxy. Tasks outlined in green, blue, and purple are (respectively) one, two or three steps below branching points in the decision tree. Table 2 gives a description of the responses that correspond to the icons shown here.

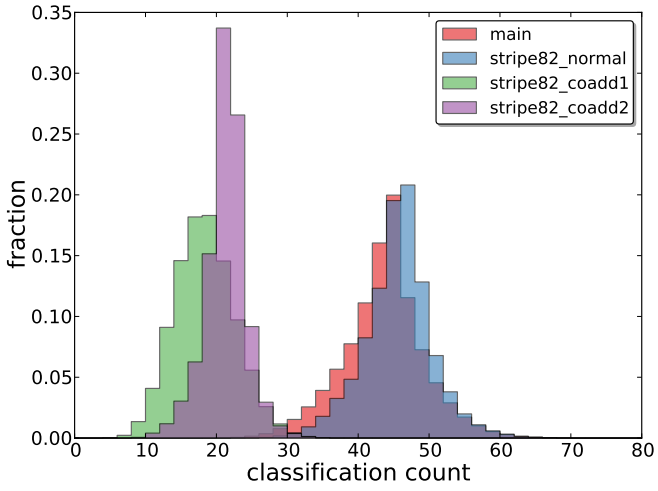

Figure 2. Distribution of the number of classifications for the sub-samples within GZ2.

$$\kappa = \frac{1}{N_r} \sum_i \kappa_i, \tag{2}$$

where N_r is the total number of possible responses for a task and

$$\kappa_i = \begin{cases} f_r & \text{if vote corresponds to this response,} \\ (1 - f_r) & \text{if vote does not correspond.} \end{cases} \tag{3}$$

For example, if a question has three possible responses, and the galaxy corresponds best to response a, then the vote fractions for responses (a, b, c) might be $(0.7, 0.2, 0.1)$.

- If an individual votes for response a, then $\kappa = (0.7 + (1 - 0.2) + (1 - 0.1))/3 = 0.8$
- If an individual votes for response b, then $\kappa = ((1 - 0.7) + 0.2 + (1 - 0.1))/3 = 0.467$
- If an individual votes for response c, then $\kappa = ((1 - 0.7) + (1 - 0.2) + 0.1)/3 = 0.4$

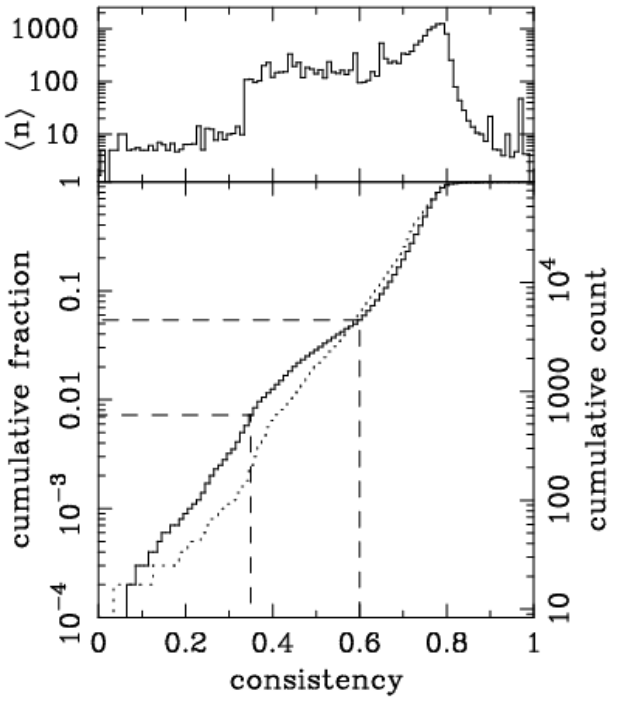

Figure 3. Distribution of the user consistency κ. Top: mean number of galaxies classified per user as a function of their consistency. Bottom: Cumulative distribution of consistency. The dotted line shows the first iteration of weighting, and the solid line the third iteration. The second iteration is not shown, but is almost identical to the third. Dashed lines indicate where the user weighting function takes values of 0.01 and 1.

Votes which agree with the majority thus have high values of consistency, whereas votes which disagree have low values.

Each user was assigned an overall consistency ($\bar{\kappa}$) by taking the mean consistency of every response. From the distribution of results for the initial iteration (Figure 3), we applied a weighting function that down-weights classifiers in the tail of low consistency.

$$w = \min\left(1.0, (\bar{\kappa}/0.6)^{8.5}\right) \quad (4)$$

For this function, $w = 1$ for $\sim 95\%$ of classifiers and $w < 0.01$ for only $\sim 1\%$ of classifiers. The vast majority of classifiers are thus treated equally; there is no up-weighting of the most consistent classifiers. The top panel of Figure 3 also shows that the lowest-weighted classifiers on average completed only a handful (< 10) of objects. This effect demonstrates either learning during classification, or the systematic loss of inconsistent people during their career as classifiers; further work on user behaviour is needed to distinguish between the two possibilities.

After computing κ, vote fractions were recalculated using the new user weights. We repeated this process a third time to ensure convergence. For each task, this produces both a weighted number of votes and a weighted vote fraction for each task. These weighted data are used exclusively hereafter when discussing votes and vote fractions; for brevity, we typically drop the term "weighted".

3.3 Classification bias

We also adjust the vote fractions for what we term *classification bias*. The overall effect of this bias is a change in observed morphology fractions as a function of redshift *independent of any true evolution in galaxy properties*, a trend also seen in the Galaxy Zoo 1 data (Bamford et al. 2009). The SDSS survey is expected to be shallow enough to justify an assumption of no evolution, and so the presumed cause is that more distant galaxies, on average, are both smaller and dimmer in the cutout images. As a result, finer morphological features are more difficult to identify. We note that this effect is not limited to crowd-sourced classifications; expert classifications must also suffer from bias to some degree, although smaller sample sizes make this difficult to quantify.

Figure 4 demonstrates the effect of classification bias for the GZ2 classification tasks. The average vote fraction for each response is shown as a function of redshift; the fraction of votes for finer morphological features (such as identification of disk galaxies, spiral structure, or galactic bars) decreases at higher redshift. The trend is strongest for the initial classification of smooth and feature/disk galaxies, but almost all tasks exhibit some level of change.

Part of the observed trends in type fractions at high redshifts is due to the nature of a magnitude-limited sample; high-redshift galaxies must be more luminous to be detected in the SDSS and are thus more likely to be giant red ellipticals. However, we see clear evidence of the classification bias even in luminosity-limited samples (between the dashed vertical lines in Figure 4). Since this bias contaminates any potential studies of galaxy demographics over the sample volume, it must be corrected to the fullest possible extent.

Bamford et al. (2009) corrected for classification bias in the GZ1 data, but only for the elliptical and combined spiral variables. Their approach was to bin the galaxies as a function of absolute magnitude (M_r), the physical Petrosian half-light radius (R_{50}), and redshift. They then computed the average elliptical-to-spiral ratio for each (M_r,R_{50}) bin in the lowest redshift slice with significant numbers of galaxies; this yields a local baseline relation which gives the (presumably) unbiased morphology as a function of the galaxies' *physical*, rather than *observed* parameters. From the local relation, they derived a correction for each (M_r,R_{50},z) bin and then adjusted the vote fractions for the individual galaxies in each bin. The validity of this approach is justified in part since debiased vote fractions result in a consistent morphology-density relation over a range of redshifts (Bamford et al. 2009). We modify and extend this technique for the GZ2 classifications as described below.

There are two major differences between the GZ1 and GZ2 data. First, GZ2 has a decision tree, rather than a single question and response for each vote. This means that all tasks, with the exception of the first, depend on responses to previous tasks in the decision tree. For example, the bar question is only asked if the user classifies a galaxy as having "features or disk" and as "not edge-on". Thus, the value of the vote fraction for this example task only addresses the total bar vote fraction *among galaxies that a user has classified as disks and are not edge-on*, and not as a function of the general population.

For a galaxy to be used in deriving a correction, we

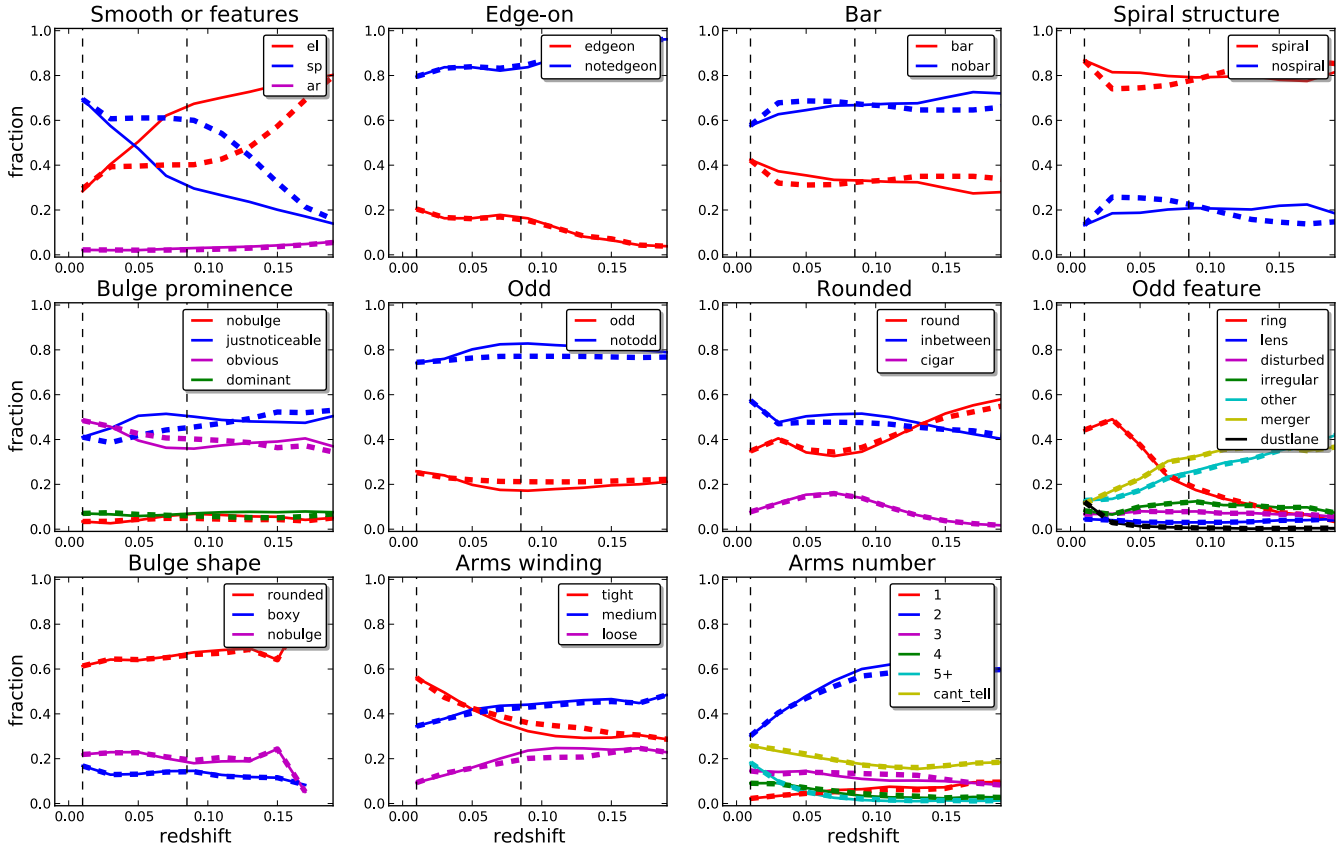

Figure 4. Type fractions as a function of redshift for the classification tasks in GZ2. Solid (thin) lines show the vote fractions, while the thick (dashed) lines show the debiased vote fractions adjusted for classification bias. This is a luminosity-limited sample for $M_r < -20.89$. The data for each task is plotted only for galaxies with enough votes to characterize the response distribution (Table 3). Vertical dashed lines show the redshift at $z = 0.01$ (the lower limit of the correction) and $z = 0.085$ (the redshift at which the absolute magnitude limit reaches the sensitivity of the SDSS).

Task	Previous tasks	Vote fraction $n_t \geqslant 10$	Vote fraction $n_t \geqslant 20$
01	–	–	–
02	01	0.227	0.430
03	01,02	0.519	0.715
04	01,02	0.519	0.715
05	01,02	0.519	0.715
06	–	–	–
07	01	0.263	0.469
08	06	0.223	0.420
09	01,02	0.326	0.602
10	01,02,04	0.402	0.619
11	01,02,04	0.402	0.619

Table 3. Thresholds for determining well-sampled galaxies in GZ2

therefore require both a minimum weighted vote fraction for the preceding response(s) and a minimum number of votes for the task in question (Table 3). While this threshold increases the number of bins with large variances, it is critical for reproducing accurate baseline measurements of individual morphologies. The correction derived from well-classified galaxies is then applied to the vote fractions for *all* galaxies in the sample.

The second major difference is that the adjustment of the GZ1 vote fractions assumed that the single task was essentially binary. Since almost every vote in GZ1 was for a response of either "elliptical" or "spiral" (either anticlockwise, clockwise, or edge-on), this ratio was employed as the sole metric. No systematic debiasing was done for the other GZ1 response options ("star/don't know" or "merger"), and the method of adjusting the vote fractions assumes that these other options do not significantly affect the classification bias for the most popular responses. This is not possible for GZ2: many tasks have more than two possible responses and represent a continuum of relative feature strength, rather than a binary choice.

Vote fractions for each galaxy are adjusted for classification bias using the following method. The method relies on the assumption that for a galaxy of a given physical brightness and size, a sample of other galaxies with similar brightnesses and sizes will (statistically) share the same average mix of morphologies. We quantify this using the ratio of vote fractions (f_i/f_j) for responses i and j. We assume that the true (that is, unbiased) ratio of likelihoods for each task (p_i/p_j) is related to the measured ratio via a single multiplicative constant $K_{j,i}$:

$$\frac{p_i}{p_j} = \frac{f_i}{f_j} \times K_{j,i}. \quad (5)$$

The unbiased likelihood for a single task can be written as:

$$p_i = \frac{1}{1/p_i}, \quad (6)$$

with the requirement that the sum of all likelihoods for a given task must be unity:

$$p_i + p_j + p_k + \dots = 1, \quad (7)$$

Multiplying (6) by the inverse of (7) yields:

$$p_i = \frac{1}{1/p_i} \times \frac{1}{p_i + p_j + p_k + \dots} \quad (8)$$

$$p_i = \frac{1}{p_i/p_i + p_j/p_i + p_k/p_i + \dots} \quad (9)$$

$$p_i = \frac{1}{\sum\limits_{j \neq i} (p_j/p_i) + 1} \quad (10)$$

$$p_i = \frac{1}{\sum\limits_{j \neq i} K_{j,i}(f_j/f_i) + 1}. \quad (11)$$

The corrections for each pair of tasks can be directly determined from the data. At the lowest sampled redshift bin, $\frac{p_i}{p_j} = \frac{f_i}{f_j}$ and $K_{j,i} = 1$. From Equation 5:

$$\left(\frac{f_i}{f_j}\right)_{z=0} = \left(\frac{f_i}{f_j}\right)_{z=z'} \times K_{j,i} \quad (12)$$

$$K_{j,i} = \frac{(f_i/f_j)_{z=z'}}{(f_i/f_j)_{z=0}} \quad (13)$$

This can be simplified if we define $C_{j,i} \equiv \log_{10}(K_{j,i})$:

$$C_{j,i} = log_{10}\left(\frac{f_i}{f_j}\right)_{z=0} - \log_{10}\left(\frac{f_i}{f_j}\right)_{z=z'}. \quad (14)$$

The correction $C_{j,i}$ for any bin is thus the difference between f_i/f_j at the desired redshift and that of a local baseline, if the ratios between vote fractions are expressed as logarithms.

The local baselines and subsequent corrections are derived from the GZ2 main sample data. Since determining the baseline ratio relies on absolute magnitude and physical size, we only use the 86% of galaxies in the main sample with spectroscopic redshifts. We also use data only from galaxies with sufficient numbers of responses to determine their morphology; this threshold is different for each task (Table 3).

The vote fractions for each task response are binned according to three parameters: the absolute magnitude M_r, the Petrosian r-band half-light radius R_{50}, and redshift z. Bins for M_r range from -24 to -16 in steps of 0.25 mag, for R_{50} from 0 to 15 kpc in steps of 0.5 kpc, and for z from 0.01 to 0.26 in steps of 0.01. These bin ranges and step sizes are chosen to maximize the parameter space covered by the bias correction. Only bins with at least 20 galaxies are considered.

Since each unique pair of responses to a question will have a different local baseline, there are $\binom{n}{2}$ correction terms for a task with n responses. For $n = 2$, this method is identical to that described in Bamford et al. (2009).

The baseline morphology ratios for the GZ2 tasks are shown in Figure 5 for the first two responses in each task. To derive a correction for bins not covered at low redshift, we attempted to fit each baseline ratio with an analytic, smoothly-varying function. The baseline ratio for the "smooth" and "features/disk" responses to Task 01 is functionally very similar to the GZ1 relation (Figure A5 in Bamford et al. 2009), as expected. This ratio can be fit with an analytic function:

$$\frac{f_j}{f_i}[R_{50}, M_r] = \frac{s_6}{1 + exp[(\alpha - M_r)/\beta]} + s_7 \quad (15)$$

where:

$$\alpha = s_2 \times \exp[-(s_1 + s_8 {R_{50}}^{s_9})] + s_3, \quad (16)$$

$$\beta = s_4 + s_5(x_0 - s_3), \quad (17)$$

where $\{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9\}$ are minimised to fit the data. The only other task that had baseline ratios well fit by an expression of this form was Task 07 (the roundedness of smooth galaxies). We adopted the same approach for this task and were able to fit the behavior of all three pairs of responses with the same functional form.

None of the other tasks are well-fit by a function of the form in Equation 15. For these, we instead adopt a simpler fit where both M_r and R_{50} vary linearly:

$$\frac{f_j}{f_i}[R_{50}, M_r] = t_1(R_{50} - t_2) + t_3(M_r - t_4) + t_5, \quad (18)$$

where $\{t_1, t_2, t_3, t_4, t_5\}$ are the parameters to be minimized. We fit Equation 18 to all other tasks where enough non-zero bins exist to get a good fit. Finally, for pairs of responses with only a few sampled bins, we instead used the difference between the local ratio and the measured ratio at higher redshift. Galaxies falling in bins that are not well-sampled are assigned a correction of $C_{i,j} = 0$ for that term; this is necessary to avoid overfitting based on only a few noisy bins.

The success of this method is good for most GZ2 tasks and responses. Figure 4 illustrates the comparison between the mean raw and debiased vote fractions as a function of redshift. The debiased results (*thick lines*) are flat over $0.01 < z < 0.085$, where $L^\star$ galaxies ($M_r \sim -20.44$; Blanton et al. 2003) are within the detection limit of the survey and there are fewer empty bins. The debiased early- and late-type fractions of 0.45 and 0.55 agree with the GZ1 type fractions derived by Bamford et al. (2009) for the same selection criteria. The bar fraction in disk galaxies is approximately 0.35, slightly higher than the value found by using thresholded GZ2 data in Masters et al. (2011).

3.4 Angular separation bias

The vote fractions also suffer from a bias which depends on the angular separation between galaxies. For some classifications, participants perceive a galaxy's morphology differently when it has a close apparent companion. Casteels et al. (2013) found that this bias is particularly strong for Task 08 ("odd features"), and in particular the "merger" classification. The mean merger vote fractions of both physically close galaxies with similar redshifts and projected pairs with very different redshifts increase strongly as a function of decreasing angular separation. This results in projected

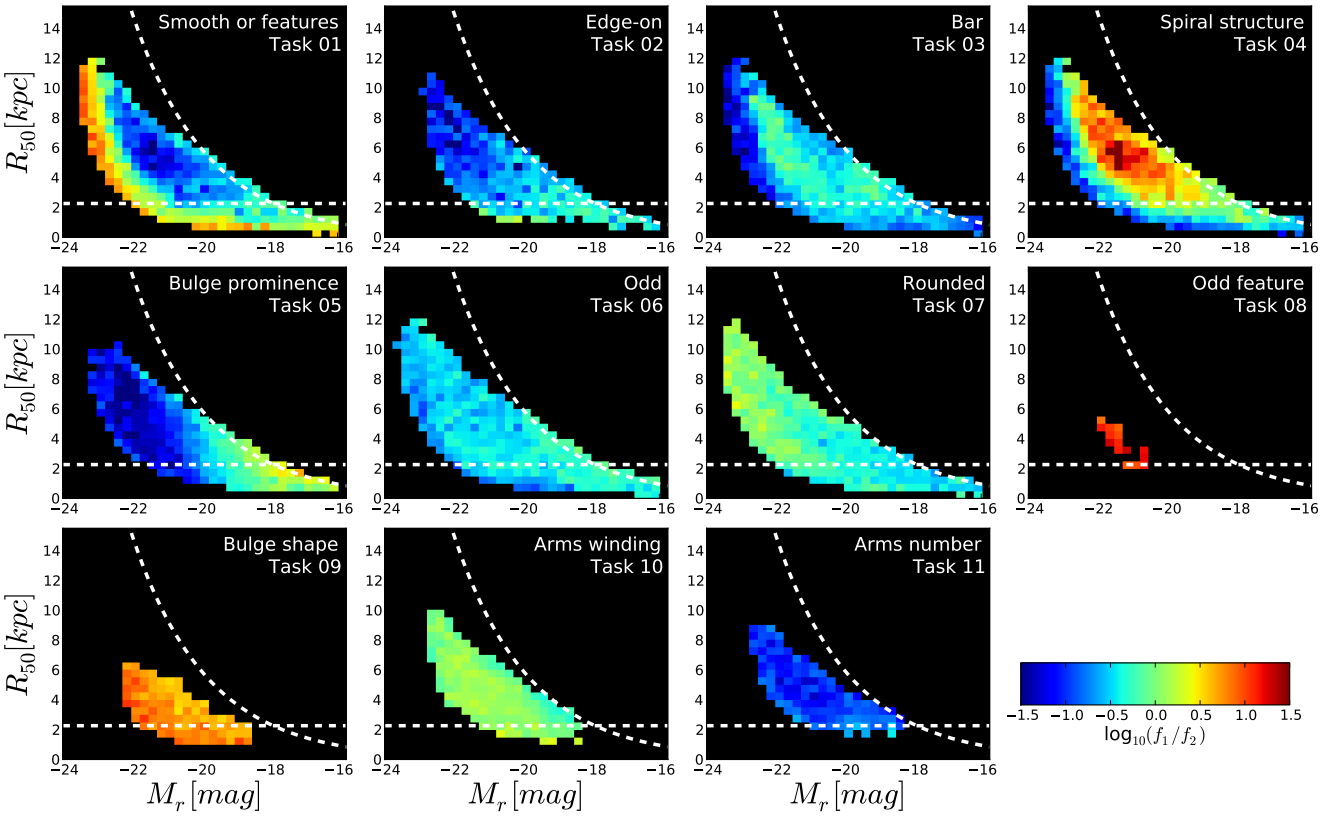

Figure 5. Local morphology ratios for GZ2 classifications; these are used to derive the corrections that adjust data for classification bias (§3.3). The ratio of the binned vote fractions is for the first two responses in the decision tree (Table 2) for each task; there may be as many as 21 such pairs per task, depending on the number of unique responses. Dashed horizontal lines give the physical scale corresponding to 1″, while the curved lines show a constant apparent surface brightness of $\mu_{50,r} = 23.0$ mag arcsec^{-2}.

pairs of non-interacting galaxies being classified as mergers. To determine an unbiased estimate of the mean probability for a given classification, Casteels et al. (2013) subtract the mean probabilities of projected pairs (with very different redshifts) from physically close pairs (with similar redshifts) for each projected separation bin. This results in a residual probability which is considered to represent the true change in morphology due to strong tidal interaction. While such a correction can be applied in a statistical way to the mean vote fractions for a given classification, applying such a correction to the individual vote fractions is not as straightforward.

The vote fractions presented in this data release have not been corrected for angular separation bias and readers using GZ2 data to study very close pairs are advised to keep in this in mind, particularly for Task 08. Fortunately, the angular separation bias is minimal for the rest of the classifications and can usually be ignored. For a detailed discussion of the angular separation bias, and how individual classifications are affected, the reader is referred to Casteels et al. (2013).

4 THE CATALOGUE

The data release for GZ2 includes the vote counts and fractions (raw, weighted, and debiased) for each task in the classification tree for each galaxy. Data for the five subsamples described below can be accessed at `http://data.galaxyzoo.org`, and will be available on CasJobs[2] in SDSS Data Release 10 after Jul 2013. Abridged portions of each data table are included in this paper (Tables 5–9).

4.1 Main sample

Table 5 contains classification data for the 243,500 galaxies in the main sample with spectroscopic redshifts. Each galaxy is identified by its unique SDSS DR7 objID, as well as its original sample designation (either original, extra or Stripe 82 normal-depth). N_{class} is the total number of volunteers who classified the galaxy, while N_{votes} gives the total number of votes summed over all classifications and all responses. For each of the 37 morphological classes, we give six parameters: the raw number of votes for that response (eg, `t01_smooth_or_features_a01_smooth_count`), the number of votes weighted for consistency (`*_weight`), the fraction of votes for the task (`*_fraction`), the vote fraction weighted for consistency (`*_weighted_fraction`), the debiased likelihood (`*_debiased`), which is the weighted vote fraction adjusted for classification bias (see Section 3.3), and a Boolean

[2] `http://skyserver.sdss3.org/casjobs/`

flag (`*_flag`) that is set if the galaxy is included in a clean, debiased sample (as described below).

Flags for each morphological parameter are determined by applying three criteria. First, we require that the vote fraction for the preceding task(s) must exceed some threshold (Table 3) to ensure that the question is well-answered. For example, to select galaxies from which a clean barred sample can be identified, we require both $p_{features/disk} \geqslant 0.430$ and $p_{notedge-on} \geqslant 0.715$. Secondly, the task must exceed a minimum number of votes (10 for Stripe 82, 20 for the main sample) in order to eliminate variance due to small-number statistics. Finally, the debiased vote fraction itself must exceed a given threshold of 0.5 for Tasks 02 and 03, and 0.8 for all other tasks. We note this is a highly conservative selection – each of the above parameters may be adjusted to provide different clean thresholds, depending on the use case for the data.

Table 6 gives the GZ2 classifications for the 42,462 main sample galaxies without spectroscopic redshifts. To compute the debiased likelihoods, we used the morphology corrections obtained for galaxies in the spectroscopic main sample. We then used SDSS photometric redshifts (Csabai et al. 2003) to derive M_r and R_{50} for each galaxy in the photometric sample and selected the appropriate correction bin. The mean error in the redshift of the photometric sample (from the SDSS photo-z) is $\Delta\ z = 0.021$ (a fractional uncertainty of 27%), compared to the spectroscopic accuracy of $\Delta\ z = 0.00016$ (0.3%). Since the size of the redshift bins in $C_{j,i}$ is 0.01, a shift of several bins can potentially produce a very large change in the debiased vote fractions.

Since the redshift can have a strong effect on classification bias, we separate galaxies with spectroscopic and photometric redshifts, and do not recommend that the debiased data be combined for analysis. For use cases where the main driver is the number of galaxies, however, it may be possible to combine the raw vote fractions for the two samples.

4.2 Stripe 82

The data for Stripe 82 is reduced separately from the main sample. This is due to the deeper magnitude limit of the samples (both normal and coadded) as well as the improved seeing in the latter. Since different image qualities can potentially affect the debiasing, all three Stripe 82 samples have their classifications separately adjusted for classification bias. The method is the same as that used for the spectroscopic main sample galaxies – the only difference is that the vote threshold for the coadded sample is lowered from 10 to 20.

Table 7 gives classifications for the Stripe 82 normal-depth images with spectroscopic redshifts. Galaxies in this table with $m_r < 17.0$ also appear in Table 5; however, the corrections for classification bias here are derived based only on Stripe 82 data, and so debiased likelihoods and flags may be slightly different. Classifications for galaxies with photometric redshifts only are not included.

Tables 8 and 9 contains classifications for the coadded images of Stripe 82 galaxies with spectroscopic redshifts. Debiased probabilities and flags are derived from separately for each coadded data set. Since both the number of galaxies and the average number of classifications per galaxy are a small fraction of that in the main sample, though, the corrections encompass a smaller range of tasks and phase space in (M_r,R_{50},z). The increased exposure time and improved seeing, however, means that the effect of classification bias is lessened at lower redshifts; the raw vote fractions may thus be more suitable for some science cases using the coadded images.

Figure 6 compares the results of the GZ2 classifications for main sample galaxies to Stripe 82, using Task 01 vote fractions as an example. The distributions of the responses for both the main sample and Stripe 82 normal-depth show the same behavior as a function of redshift. This applies both when using thresholded vote fractions and the raw likelihoods. The type fractions for the coadded data, however, are significantly different – classifications of coadded images have a significant increase in the fraction of responses for "features or disk". This results in a large increase in the number of unclassified galaxies (and subsequent decrease in the number of clean smooth galaxies) when using thresholds, and a similar shift of vote fractions from smooth to feature/disk when using the raw likelihoods.

This significant difference in the classification behavior demonstrates why the main sample corrections cannot be applied to the coadded images. The likely cause is that the coadded data allows classifiers to better distinguish faint features and/or disks, due to both improved seeing (from 1.4″ to 1.1″; Annis et al. 2011) and higher signal-to-noise ratio.

Comparing the coadd1 and coadd2 images, the data show no systematic differences between classifications for the majority of the GZ2 tasks. Figure 7 shows the difference between the two vote fractions ($\Delta_{coadd} = p_{coadd1} - p_{coadd2}$) for four examples. If the mean value of Δ_{coadd} for a response was non-zero, that would indicate a systematic bias in classification due to the image processing. In GZ2, 33/37 tasks have $|\Delta_{coadd}| < 0.05$ (for galaxies with at least 10 responses to the task), with variations in the mean scattered on both sides of Δ_{coadd}.

The biggest systematic difference is for the response to Task 05 (bulge prominence) of the bulge being "just noticeable". The mean fraction in coadd2 data is 35% higher than that in coadd1 data. This effect is opposite (but not equal) to that for an "obvious" bulge, for which the coadd1 data is 13% higher; this may indicate a general shift in votes toward a more prominent bulge. A similar but smaller effect is seen in classification of bulge shapes for edge-on disks (Task 09), where votes for "no bulge" in coadd1 data go to "rounded bulge" in coadd2. The specific cause for these effects as it relates to the image quality is not investigated further in this paper.

For most morphological questions, the two versions of coadded images showed no significant difference. While either set of coadded data can likely be used for science, we recommend using coadd2 if picking between them. The overall consistency indicates that the votes for both could potentially be combined and treated as a single data set; this could be useful for increasing the classification accuracy for deeper responses (such as arms number or arms winding) within the GZ2 tree.

4.3 Using the classifications

Since GZ2 is intended to be a dataset for use by the community, we present two examples of how classifications can be

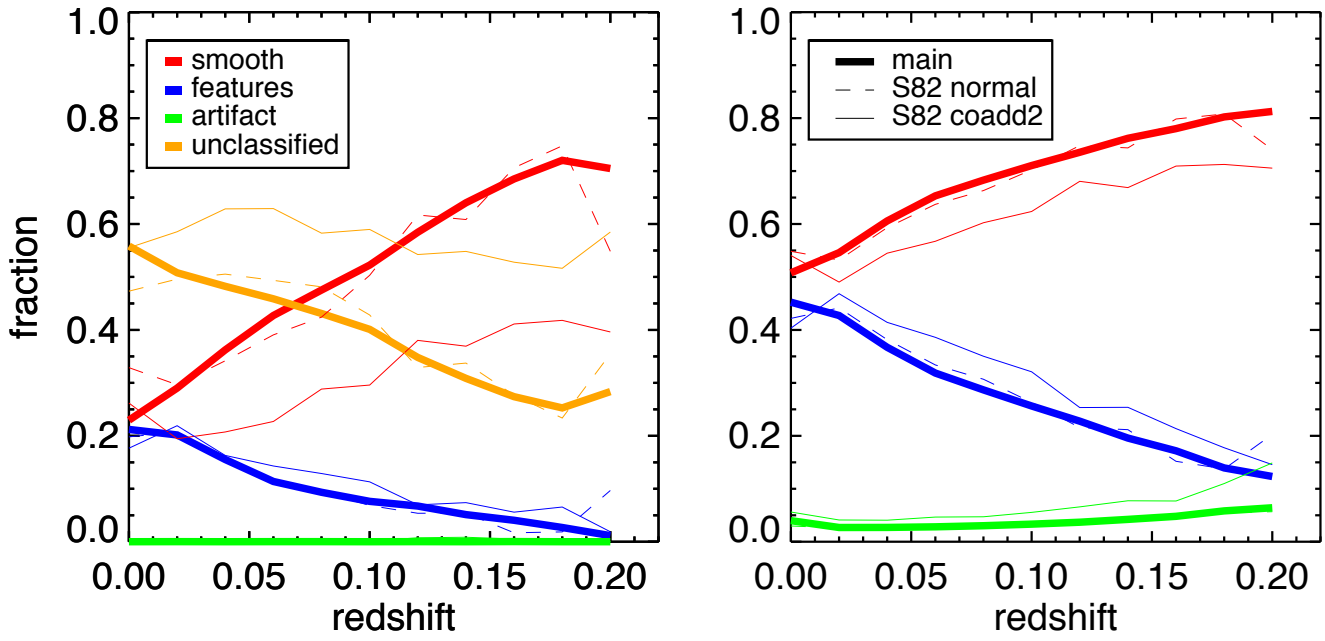

Figure 6. GZ2 vote fractions for Task 01 (*smooth, features/disk, or star/artifact?*) as a function of spectroscopic redshift. *Left:* fraction of galaxies for which the GZ2 vote fraction exceeded 0.8. Galaxies which did not have a response above 0.8 are labeled "unclassified". *Right:* mean GZ2 vote fractions weighted by the total number of responses per galaxy. Data are shown for the GZ2 original + extra (thick solid), Stripe 82 normal-depth (thin dotted), and Stripe 82 co-add depth (thin solid) samples with a magnitude limit of $m_r < 17.0$.

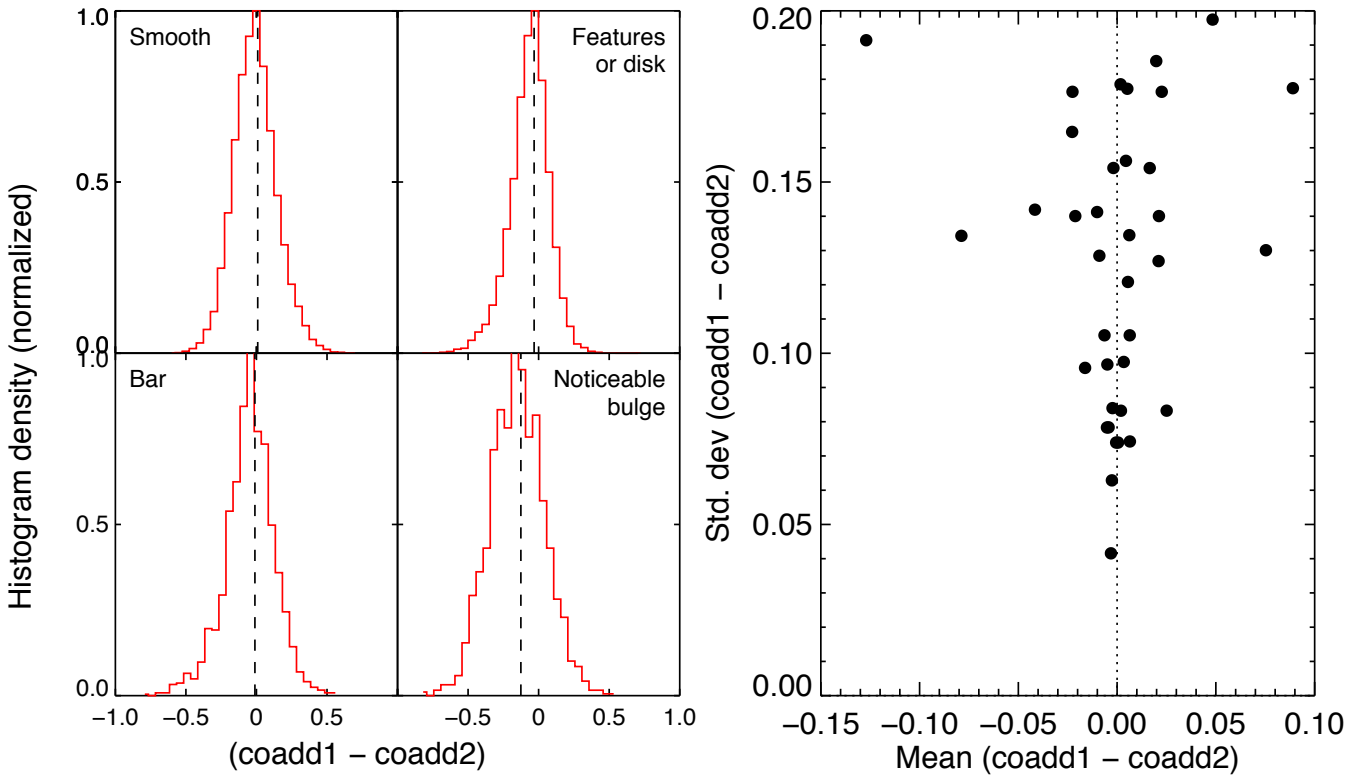

Figure 7. Comparison of GZ2 classifications for the coadded images of Stripe 82. Left: Distribution of the difference in vote fractions for galaxies that appear in both the coadd1 and coadd2 samples. Four example tasks are shown, including only galaxies with at least 10 responses per task. The dashed line shows the median of each distribution, showing a considerable amounts of scatter for individual responses. "Noticeable bulge" was the only response in GZ2 for which the mean $|\Delta_{coadd}| > 0.1$. Right: mean values of the difference in the vote fractions for every response in the GZ2 tree.

used and selected from the catalogue. These will depend on the individual science cases, and additional cuts (eg, making a mass or volume-limited sample) may be required to define the parameters more appropriately.

The first use case suggested for the GZ2 data is the selection of pure samples matching a specific morphology category. This is appropriate for when some finite number of objects with clear morphological classifications is required (perhaps for individual study or an observing proposal), but there is no requirement to have a complete sample. An example would be the selection of three-armed spirals. The simplest way is to search for galaxies in the GZ2 spectroscopic main sample (Table 5) with `t11_arms_number_a33_3_flag`= 1, which returns 308 examples. Inspection of the flagged images shows that they are all in fact disk galaxies with three spiral arms, with no object that is a clear false positive.

Alternatively, users of the catalogue can set their own thresholds for the morphology tasks if they wish to change the strength of the cut. This flag is currently set via the combination of $p_{features/disk} > 0.430$, $p_{edge-on,no} > 0.715$, $p_{spiral,yes} > 0.619$, $N_{spiral,yes} > 20$, and $p_{3\ arms} > 0.8$ (Table 3). These cuts are generally regarded as conservative, and more genuine three-armed spirals can be discovered by, for example, lowering the threshold on $p_{3\ arms}$. If the number of objects returned by such a query is of a manageable size, we suggest that the images be individually examined – this is the only way to differentiate spirals with true radial symmetry from 2-armed spirals with an additional tidal tail, for example. Galaxy Zoo 2 papers that employ similar methods of selecting specific morphologies include Masters et al. (2011), Kaviraj et al. (2012), and Simmons et al. (2013).

The second common use case for the morphologies is the direct use of the likelihoods. While thresholds on the likelihoods are appropriate for some studies, most classifications do not exceed $p > 0.8$ for any available response, especially when more than two are available. These intermediate classifications are a combination of genuine physical attributes of the galaxy (vote fractions of $p_{smooth} = 0.5, p_{features/disk} = 0.5$ may accurately represent a galaxy with strong bulge and disk components) in addition to limitations in accuracy from the image quality and variance among individual classifiers. The problem is that thresholding only samples a small portion of the vote distributions.

In order to use data for the *entire* sample, the debiased likelihoods for each response can be treated as probabilistic weights. As an example, consider the type fractions shown in Figure 6. The left hand side shows the average fraction of morphological classes at each redshift only defining a "class" as exceeding some vote fraction threshold; as a result, more than half the galaxies are left "unclassified", with no strong majority. The panel on the right uses the likelihoods themselves directly. A galaxy with $p_{smooth} = 0.6$, $p_{features/disk} = 0.3$, and $p_{star/artifact} = 0.1$ contributes 0.6 of a "vote" to smooth, 0.3 to features/disk, and 0.1 to star/artifact. This approach is generally suitable for studying morphology dependence on global variables, such as environment or color. Further examples of using the likelihoods as weights can be found in Bamford et al. (2009) and Skibba et al. (2012).

GZ2	GZ1		HC11		NA10		EFIGI	
	N	%	*N*	%	*N*	%	*N*	%
early-type	79214	86.2	26732	82.1	1995	96.7	214	84.6
late-type	26314	97.9	79277	88.6	5481	94.9	1675	98.2
bar	–	–	–	–	651	94.9	238	98.7
ring	–	–	–	–	438	91.6	110	83.6
merger	526	63.3	–	–	43	100.	6	100.

Table 4. Comparison of the agreement in morphology between the GZ2 main sample and other catalogs. For each category, a galaxy is considered to "agree" if it has a likelihood of at least 0.8 ("clean") in GZ2 and at least 0.5 ("majority") in the other catalog (or for NA10, inclusion in the relevant category). Table gives both the total number of GZ2 galaxies and the fraction that agree in the other catalog.

5 COMPARISON OF GZ2 TO OTHER CLASSIFICATION METHODS

To assess both the scope and potential accuracy of the GZ2 classifications, we have compared our results to four recent morphological catalogues (including the previous version of Galaxy Zoo). All four catalogs contain classifications based on optical SDSS images and have significant overlaps with the galaxies in GZ2.

- Galaxy Zoo 1 (Lintott et al. 2011)
- Nair & Abraham (2010)
- EFIGI (Baillard et al. 2011)
- Huertas-Company et al. (2011)

5.1 Galaxy Zoo 1 and Galaxy Zoo 2

The images from in GZ2 are a subset of GZ1, with 248,883 galaxies included in both catalogs. Task 01 in GZ2 is broadly similar to the interface of GZ1, with some modifications. GZ1 had six possible responses for its task: "elliptical", "clockwise spiral", "anticlockwise spiral", "other spiral", "merger" and "star/don't know". We compared vote fractions for the GZ1 "elliptical" to GZ2 "smooth", and combined responses for all GZ1 spiral categories to the GZ2 "features or disk".

The matched GZ1-GZ2 catalogue contains 34,480 galaxies identified as ellipticals based on their debiased GZ1 likelihoods (Lintott et al. 2011). Of those, 89.0% had GZ2 vote fractions for "smooth" greater than 0.8 and 99.9% greater than 0.5. Using the GZ2 debiased likelihoods, 50.4% of galaxies have vote fractions exceeding 0.8 in both samples, while 97.6% have vote fractions exceeding 0.5.

The matched catalog also contains 83,956 galaxies identified as spirals by GZ1. The agreement with the "features or disk" response in GZ2, however, is significantly lower than that of ellipticals. Only 31.6% of the GZ1 clean spirals had GZ2 vote fractions greater than 0.8, with 59.2% greater than 0.5. The GZ2 debiased likelihoods for the same galaxies only match at 38.1% (for 0.8) and 78.2% (for 0.5).

Figure 8 shows the difference between the vote fractions for the spiral classifications in GZ1 and features/disk classifications in GZ2 for all galaxies that appear in both catalogues. The vote fractions show a tight correlation at both very low and very high values of the GZ1 vote fraction for combined spiral (f_{sp}), indicating that both projects agree on the strongest spirals (and corresponding ellipticals). At

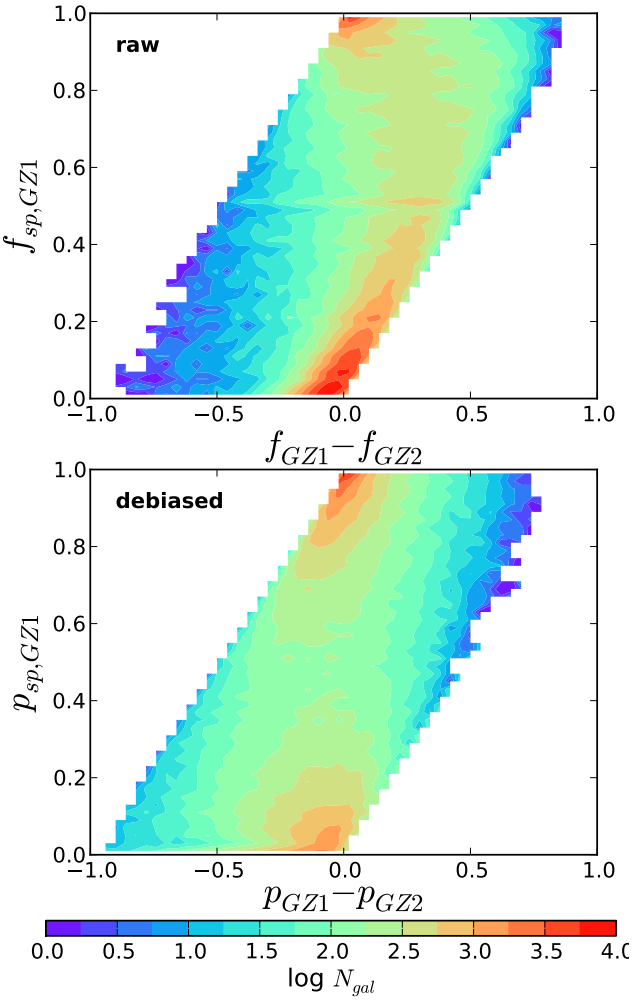

Figure 8. Comparison of spiral galaxies using classifications for "combined spiral" (GZ1) and "features or disk" (GZ2). Left: raw vote fractions. At intermediate values ($f_{sp} \sim 0.5$), GZ1 classifiers are more likely to identify galaxies as spiral compared to GZ2. Right: debiased vote fractions. At intermediate values, GZ1 and GZ2 classifications are consistent with each other; however, there is an increased scatter in the vote fractions near $p_{sp} \simeq 0$ and $p_{sp} \simeq 1$.

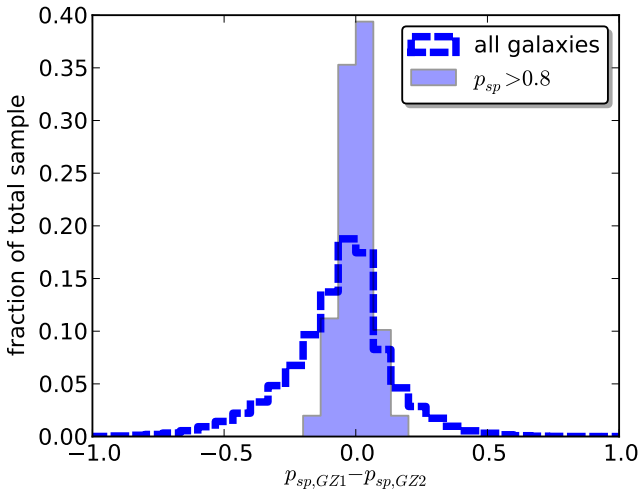

Figure 9. Comparison of the spiral feature vote fractions for objects in Galaxy Zoo 1 (GZ1) and Galaxy Zoo 2 (GZ2). The dashed line shows the difference between $p_{combined\ spiral}$ for GZ1 and $p_{features\ or\ disk}$ for GZ2 for the 240,140 galaxies in both samples. The filled histogram shows the same metric for the 57,994 galaxies classified as "clean" spirals in both GZ1 and GZ2.

intermediate (0.2–0.8) values of f_{sp}, however, GZ1 has vote fractions that are consistently higher than those in GZ2, differing by up to 0.25. When using debiased likelihoods in place of the vote fractions, this effect decreases dramatically; however, the tightness of the correlation correspondingly drops at low and high p_{sp}.

Based on the vote fractions, GZ2 is significantly more conservative than GZ1 at identifying spiral structure. A possible explanation is that this is a bias from classifiers who are anticipating subsequent questions about the details of any visible structures. An experienced classifier, for example, would know that selecting "features or disk" is followed by additional questions, none of offer options for an uncertain classification. If the classifier is less confident in identifying a feature, it is possible they would avoid this by clicking "smooth" instead. This is a hypothesis; there is no direct evidence from the data suggesting that this has taken place, but suggest it as one possibility for explaining the discrepancy in otherwise similar classification tasks.

Since the GZ1 vote fractions were specifically directed to galaxies with spiral arms, we also compared GZ1 to the results of Task 04 in GZ2, which specifically asks about spiral structure in galaxies already identified as disks that are not edge-on. The agreement with the GZ1 clean spirals is higher than for Task 01, but still well short of that for smooth/elliptical galaxies. Only 63.6% of galaxies have Task 04 vote fractions greater than 0.5 for the GZ1 clean spirals, rising to 66.8% when using the debiased data.

Figure 9 shows the distribution of the difference between the vote fractions for GZ1 and GZ2, using the debiased likelihoods for combined spirals for GZ1 and "features or disk" galaxies in GZ2. There is a slight skew in the distribution, indicating that a galaxy is more likely to be identified as a spiral in GZ2 compared to GZ1.

Galaxies are significantly more likely to be identified as a spiral in GZ1 than in GZ2. When restricted only to galaxies in the joint clean samples ($p > 0.8$), the spread is greatly reduced and the distribution is centered around a difference of zero. The debiased vote fractions show a similar spread when comparing GZ1 and GZ2 classifications, although the skew toward spirals in GZ1 is largely removed. When using only clean galaxies and the debiased vote fractions, galaxies are more likely to be identified as spirals in GZ2.

The GZ1 interface did have one option that did not classify either early- or late-type galaxies, but rather mergers. This was a rare response in the GZ1 data, comprising less than a percent of the total type fraction at all redshifts (Bamford et al. 2009). Darg et al. (2010) found that a vote fraction of $f_{mg} > 0.6$ robustly identified merging systems in GZ1. Of the 1632 such systems classified in GZ2, more than 99% were identified as "odd" galaxies, and 77.7% had $p_{mg} > 0.5$ in GZ2. This is partly due to early-stage merging spirals avoiding the "merger" classification, with only late-

state mergers with extremely disturbed morphologies recording high vote fractions for the merger question.

In addition to the angular separation bias discussed in section 3.4, classifications in Task 08 ("odd feature") also suffer from crosstalk between the possible responses. This is a result of more than one response being applicable for some galaxies, which forces the participant to choose the one they consider most relevant. For example, a merging galaxy may display a strong dust lane, be highly irregular in shape, and have a disturbed appearance. While "merger", "dust lane", "irregular" and "disturbed" are all possible classifications, the participant will usually choose the "merger" classification and information about the other morphological features is lost. For close pairs, this crosstalk is a function of angular separation – the fraction of galaxies classified as mergers increases with decreasing separation, while the other "odd feature" classifications lose votes correspondingly (Casteels et al. 2013).

For the other GZ2 tasks the classifications are mutually exclusive and only one response per task is applicable. Crosstalk is also present in the GZ1 classifications, specifically the merger class. In the latest incarnation of Galaxy Zoo[3] it is possible to select multiple classifications from the "odd feature" task so that crosstalk can be avoided.

GZ1 and GZ2 share a similar interface, and have nearly 250,000 galaxies classified in both samples. The separation of early and late-type galaxies from the two projects is mostly consistent, especially for high-confidence ($p > 0.8$) galaxies. GZ2 classifications are more conservative than GZ1 at identifying spiral structure for intermediate vote fractions. Mergers identified in GZ1 appear at a very high rate in GZ2 as "odd" galaxies, although classification as a merger is complicated by cross-talk between other GZ2 responses to Task 08.

5.2 Expert classifications

The standard for detailed morphological classifications for many years has come from identifications of individual experts. We compare the GZ2 classifications to two SDSS morphological catalogs generated by small groups of professional astronomers: Nair & Abraham (2010, hereafter NA10) and EFIGI (Baillard et al. 2011). The fact that GZ2 and both expert catalogs used data from the same survey allows for direct comparison of the classification methods.

The catalogue of NA10 is based on images of 14,034 galaxies from SDSS DR4. Galaxies were selected from a redshift range of $0.01 < z < 0.1$, with an extinction-corrected apparent magnitude limit of $g < 16$. In comparison, the GZ2 sample is deeper ($m_r < 17$), spans a wider redshift range ($0.0005 < z < 0.25$), and contains a more recent data release (DR7). 12,480 galaxies were classified in both GZ2 and NA10 – this comprises nearly all (89.9%) of the NA10 catalogue, but only 4.5% of GZ2.

NA10 is based on visual classifications of monochrome g-band images by a single astronomer (P. Nair). The data include RC3 T-types (a numerical index of a galaxy's stage along the Hubble sequence; de Vaucouleurs et al. 1991) as well as various "fine structure" morphological features in each galaxy, including bars, rings, lenses, pairs, interactions, and tails. The NA10 data does not contain information on the likelihood or uncertainty associated with morphological features, although it does measure some features by their relative strengths (dividing barred galaxies into strong, medium, and weak classes, for example).

EFIGI consists of classifications of 4,458 galaxies, a subset of the RC3 catalogue with 5-colour imaging in SDSS DR4. There was no firm redshift or volume limit, but almost all galaxies in EFIGI are at $0.0001 < z < 0.08$. 3,411 galaxies are in both EFIGI and GZ2. This constitutes 77% of EFIGI and 1.2% of the GZ2 sample. Classifications on composite *gri* images were performed by a group of 11 professional astronomers, each of whom classified a subset of 445 galaxies. A training set of 100 galaxies was also completed by all 11 astronomers to adjust for biases among individual classifiers.

EFIGI contains two types of morphological classification: T-types and attributes. T-types were assigned using a slightly modified version of the RC3 Hubble classifications. Peculiar galaxies were not considered a separate stage, and ellipticals were subdivided into various types: compact, elongated (standard elliptical), cD (giant elliptical), and dwarf spheroidals. The remaining morphological information, dubbed "attributes", is divided into six groups:

- appearance: `inclination/elongation`
- environment: `multiplicity, contamination`
- bulge: `B/T ratio`
- spiral arms: `arm strength, arm curvature, rotation`
- texture: `visible dust, dust dispersion, flocculence, hot spots`
- dynamics: `bar length, inner ring, outer ring, pseudo-ring, perturbation`

EFIGI attributes were measured on a five-step scale from 0 to 1 (0, 0.25, 0.50, 0.75, 1). For some attributes (eg, arm strength, rings), the scale is set by the fraction of the flux contribution of the feature relative to that of the entire galaxy. For others (eg, inclination or multiplicity), it ranges between the extrema of possible values. A 70% confidence interval is estimated by setting lower and upper limits on the same five-point scale.

The EFIGI and NA10 catalogues were compared in detail by Baillard et al. (2011). Classifications on the T-type scales showed a high level of agreement; EFIGI lenticular and early spirals have slightly later average classifications in NA10, while later EFIGI galaxies have slightly earlier NA10 T-types. EFIGI has a major fraction of galaxies with slight-to-moderate perturbations with no interaction flags set in the NA10 catalogue, indicating that NA10 is less sensitive toward more benign features (eg, spiral arm asymmetry). The bar length scale is consistent between the two samples; good agreement is also found for ring classifications.

We compare the GZ2 data to the expert catalogues for data covered by both classification schemes, focusing on bars, rings, and interacting galaxies. We also examine the relationship between the expert-determined T-types and the decision tree in GZ2.

[3] `www.galaxyzoo.org`

5.2.1 Bars

NA10 detect 2537 barred galaxies, 18% of their total. For objects with T-types later than E/S0, this rises to 25% of the sample. This is consistent with the bar fraction from (Masters et al. 2011) for disky, not edge-on galaxies from early GZ2 data (29%). Of the objects NA10 identify as barred galaxies, 2348 (93%) are objects in GZ2.

To analyse the overlap between bars detected in NA10 and GZ2, we restrict the comparison sample to galaxies in GZ2 identified as possessing disks and being "not edge-on" (NEDs). NED galaxies are selected from the GZ2 data as having $p_{features/disk} > 0.430, p_{not\ edgeon} >= 0.715$, and $N_{not\ edgeon} \geqslant 20$ (Table 3). This restricts the sample of overlapping galaxies in NA10 to 5,526 NEDs. The "not edge-on" cut is similar to a restriction on inclination angle of $\lesssim 70°$, based on the average axial ratio from the SDSS exponential profile fits to the galaxies.

Bars in NA10 are classified according to either bar strength (weak, intermediate, strong) or by other morphological features (ansae, peanuts, or nuclear bar). A galaxy may in rare cases have both a disk-scale (strong, intermediate, or weak) and a nuclear bar. Figure 10 (*top left*) shows that the GZ2 average vote fraction for bars closely agrees with the NA10 fraction of barred galaxies for each GZ2 bin. The two quantities are not identical; the x-axis plots *individual classifications* of galaxies with varying vote fractions for the presence of a bar. The y-axis shows the ratio of barred to unbarred galaxies in NA10. The data have a correlation coefficient of $\rho = 0.984$, and closely follow a linear relationship. For this task, the aggregate votes of volunteers closely reproduce overall trends in expert classification.

The top right panel of Figure 10 shows the distribution of GZ2 bar votes by simply splitting the NA10 sample in two: galaxies without a bar and galaxies with a bar (of any kind). Both samples show a strong trend toward extrema, with the peak near zero for non-barred galaxies indicating that GZ2 classifiers are very consistent at identifying unbarred disk galaxies. Possession of a bar is less straightforward; while the frequency of NA10 bars does increase with GZ2 fraction, 39% of barred galaxies from NA10 have a GZ2 vote fraction < 0.5. Conversely, only 6% of non-barred NA10 galaxies have GZ2 bar vote fractions above 0.5.

The bottom left panel of Figure 10 shows the distribution of GZ2 vote fraction split by bar strength from NA10. The distribution for all bars is the same as shown in the top right, increasing with GZ2 vote fraction. There is a clear difference in the GZ2 classifications as a function of NA10 bar strength; all three are statistically highly distinct from each other and from the overall barred sample, according to a two-sided K-S test. The majority of both the strong and intermediate barred population have high GZ2 vote fractions, with 78% of strong bars and 40% of intermediate bars above a bar vote fraction of 0.8. Those numbers increase to 94% and 80%, respectively, if the majority criterion of 0.5 for the GZ2 vote fraction is used (Masters et al. 2011). Only 9% of weakly-barred galaxies have GZ2 vote fractions above 0.8, and 32% have vote fractions above 0.5.

The lack of sensitivity to weak bars from NA10 may also be related to the design of the GZ2 interface. When asked if a bar is present, the image shown in the web interface is an icon with two examples of a barred galaxy (Figure 1). The

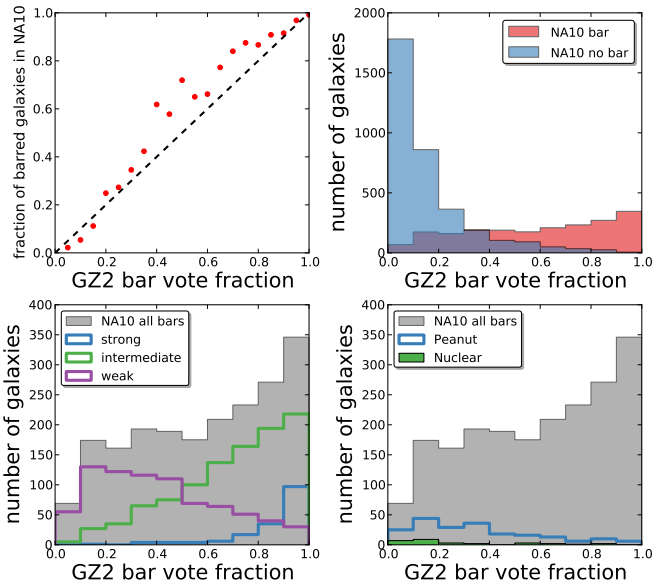

Figure 10. Classifications for galactic bars in GZ2 and NA10. Data are for the 5,526 galaxies in both samples classified by GZ2 as not-edge-on disks and with $\geqslant 20$ bar classifications. Top left: mean bar vote fraction per galaxy in GZ2 vs. the ratio of barred to all galaxies in NA10. Dashed line shows the one-to-one relationship. Top right: distribution of the GZ2 debiased bar vote fraction, separated by NA10 classifications. Bottom left: distribution of GZ2 bar vote fraction for the three disk-scale bar categories of NA10. Bottom right: distribution of GZ2 bar vote fraction for peanut and nuclear bars from NA10.

example picture of a disk galaxy has the bar extending across the disk's full diameter, fitting the typical definition of a strong bar. With this as the only example (and no continuum of options between the two choices), GZ2 participants may not have looked for bars shorter than the disk diameter, or been less confident in voting for "yes" if they did see them. Results from Hoyle et al. (2011) show that classifiers are fully capable of identifying weak bars in other contexts.

Ansae, peanuts, and nuclear bars as identified by NA10 do not correlate strongly with the GZ2 bar parameter. In fact, the median bar vote fraction for peanuts and nuclear bars (no ansae appear in the NED sample) is only $p_{bar} = 0.29$. Nuclear bars are the only feature that overlaps with the NA10 bar strength classifications; out of 283 nuclear bars, 3 galaxies also have strong bars, 44 have intermediate bars, and 166 have weak bars.

The EFIGI bar length attribute is measured with respect to D_{25}, the decimal logarithm of the mean isophote diameter at a surface brightness of $\mu_B = 25$ mag arcsec^{-2}. A value of 1.0 (the strongest bar) extends more than half the length of D_{25}, while the median value of 0.5 would be about one-third the length of D_{25}. The overall fraction of barred galaxies in EFIGI is 42% (1439/3354); this is essentially unchanged if only NEDs are considered (915/2099 = 44%). This is significantly higher than the mean bar fraction of Masters et al. (2011), at 29.5%, but consistent with results using automated ellipse-fitting techniques (Barazza, Jogee & Marinova 2008; Aguerri, Méndez-Abreu & Corsini 2009).

The higher fraction in EFIGI is due to the contributions

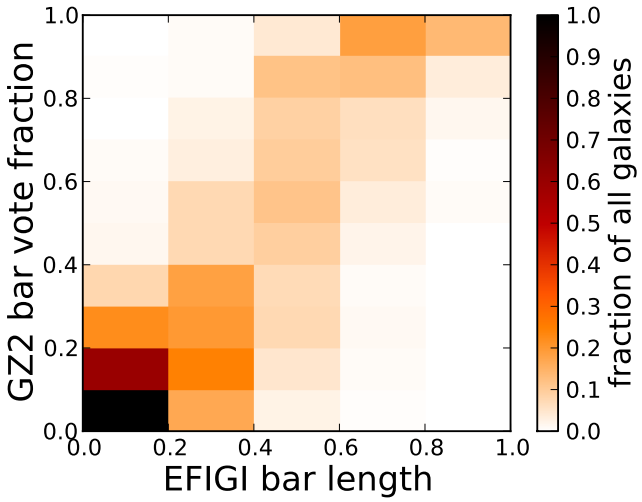

Figure 11. EFIGI bar length classifications compared to their GZ2 vote fractions for the presence of a bar. Data are for the 2,232 NED galaxies in both EFIGI and GZ2 with at least 10 bar classifications.

of galaxies with bar length attributes of 0.25, the majority of which have GZ2 vote fractions below 0.5. If only EFIGI galaxies at 0.5 and above are considered to be barred, then the bar fraction falls to 17%. Only some of the galaxies in the 0.25 EFIGI bin are being classified by the GZ2 users as barred, however, Baillard et al. (2011) defines these as a "barely visible" bars.

There is a strong correlation between the GZ2 bar vote fractions and the attribute strength from EFIGI (Figure 11). 65% of galaxies in both EFIGI and GZ2 sample have no strong evidence for a bar ($p_{bar} < 0.3$); of those, 77% had EFIGI bar attributes of zero and 94% had 0.25 or less. For higher values of the GZ2 vote fraction for bars, the EFIGI attribute is slightly lower; the largest number of galaxies with GZ2 $p_{bar} > 0.8$ have EFIGI values of 0.75. The correlation coefficient between the variables is 0.51; if only NEDs are considered, this increases to 0.75.

Using the same criteria from Table 3, there are 1,543 NEDs with an EFIGI classification. Barred galaxies as identified by GZ2 ($p_{bar} \geqslant 0.3$) agree very well with EFIGI; less than 5% of GZ2 barred galaxies have EFIGI attributes of 0, with a mean value of 0.56. This could indicate a selection preference toward medium-length bars (one-third to one-half of D_{25}), or could genuinely reflect the fact that medium bars are the most common length in disk galaxies.

Data from both NA10 and EFIGI can be used to quantify a threshold to identify barred galaxies in GZ2 data. The fraction of non-barred galaxies in NEDs as identified by both expert catalogs drops to less than 5% at a GZ2 vote fraction $p_{bar} = 0.3$, which we suggest as the preferred selection criteria. This threshold may be changed depending on the specific science needs, but offers a useful trade-off between inclusion of nearly all (97% from NA10) strong and intermediate bars and most (75%) of the weak bars. This is a slightly more inclusive threshold than the $f \geqslant 0.5$ used by Masters et al. (2011).

5.2.2 Rings

NA10 classify three types of ring galaxies based on criteria from Buta & Combes (1996): inner rings (between the bulge and disk), outer rings (external to the spiral arms), and nuclear rings (lying in the bulge region). In GZ2, rings can be identified only if the user selects "yes" for the question *"Anything odd?"* Since the "odd feature" task has seven responses, of which a user can select only one, any galaxies with multiple "odd" features will have votes split among the features, with only one option achieving a plurality. While this means that some galaxies with rings may have low vote fractions in the GZ2 classifications, those with high vote fractions are typically strong and distinct.

In the NA10 catalogue, 18.2% of all galaxies (30% of disks) have a ring. Of those, 10% are nuclear rings, 74% are inner rings, and 32% are outer rings (the sum is more than 100% since one-third of ringed galaxies have multiple rings flagged). We examined the match of NA10 ring classifications for the 7,245 NEDs appearing in both samples. No cut is applied to the vote fraction for the "anything odd" question; even a comparatively low cut of $p_{odd} > 0.2$ eliminates roughly 40% of the ringed galaxies identified in NA10. We therefore analysed the GZ2 ring vote fraction for the entire sample of NEDs, with the caveat that some 13% of these classifications are the result of three or fewer votes.

Figure 12 shows the distribution of the GZ2 ring vote fraction (p_{ring}) in the NED sample, split by the identification of a ring in NA10. While there is a marked increase in the fraction of ringed galaxies at $p_{ring} > 0.5$, more than a third of these galaxies are identified by NA10 as ringless. The agreement is significantly better if a limit is placed on the number of votes. Setting $N_{ring} > 5$, for example, increases the agreement to $\sim 75\%$.

The distribution of p_{ring} is strongly affected by the ring type in NA10. Among galaxies that NA10 identifies as rings for which GZ2 strongly disagrees ($p_{ring} < 0.5$), the majority are classified as inner rings. There are 308 ringed galaxies from NA10 that have no ring votes at all in GZ2; 84% of these are inner rings. For galaxies on which the NA10 and GZ2 ring classifications agree, the percentage of outer ring galaxies is much higher. In the absence of specific instructions on different types of ring (the icon for this response in Figure 1 does not indicate the size of the disk relative to the ring), classifiers seem much more likely to identify outer rings in GZ2. The flat distribution of p_{ring} for nuclear rings indicates that there is also no strong correlation between GZ2 classifications and ring structures in the bulge.

Most galaxies with $p_{ring} > 0.5$ are classified as outer rings in NA10, especially if constraints on N_{odd} and/or p_{odd} are added. Part of the reason for the remaining disagreements may relate to the placement of the ring classification in GZ2 at the end of the tree, and only as a result of the user identifying something "odd". Without having seen examples of ring galaxies (especially as their structures connect to spiral arms), users may have been less likely to characterize the galaxy as odd and thus do not address the ring question.

In EFIGI, classification of rings is split into inner, outer, and pseudo rings. Both outer and pseudo ring galaxies show reasonably strong correlations with GZ2 ring classifications, with a mean ring vote fraction of 0.69 for outer ring galaxies and 0.71 for pseudo-ring galaxies. The mean GZ2 ring vote

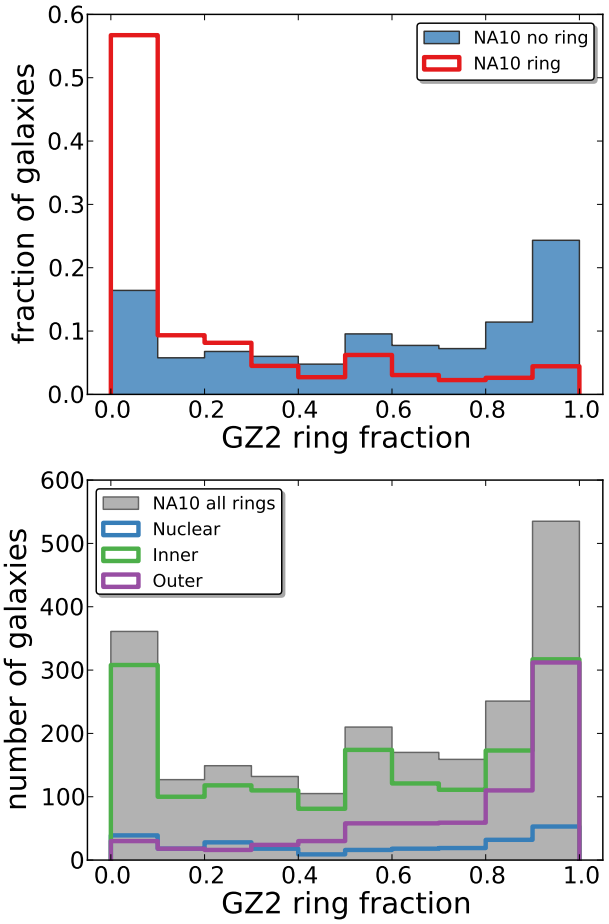

Figure 12. Ring classifications in GZ2 and NA10. Data are for the 7,245 NED galaxies in both samples. Top: GZ2 vote fraction for rings (N_{ring}/N_{odd}) for all galaxies, split by their NA10 ring identifications. Bottom: GZ2 ring vote fraction for all rings identified by NA10, separated as a function of ring type.

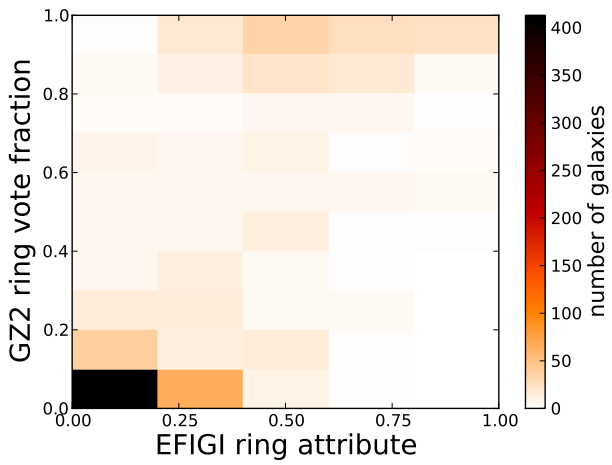

Figure 13. EFIGI ring classifications compared to their GZ2 ring vote fractions. The EFIGI data is the strongest attribute among the combined inner, outer, and pseudo-ring categories. Data are for the 1,080 galaxies in both EFIGI and GZ2 with at least 10 responses to Task 08 (odd feature).

fraction for inner rings is only 0.41. For galaxies in both EFIGI and GZ2, a high GZ2 ring vote fraction agrees significantly with the expert classification of a ring. 89% of galaxies with $p_{ring} > 0.5$ and having at least 10 votes for "Anything odd?" were classified as rings in EFIGI.

Figure 13 shows a moderate correlation between the EFIGI ring attributes and the GZ2 ring vote fractions. The relationship is dominated by galaxies for which the methods agree strongly on either no ring or a ring with high contributions to the total galaxy flux. For intermediate (between 0.25 and 0.75) values of the EFIGI ring attribute, the GZ2 vote fraction has relatively little predictive power.

5.2.3 Mergers and interacting galaxies

Galaxies in GZ2 are classified as mergers under the task *"Anything odd?"* NA10 classify possible mergers in two methods: both from pairs of objects and from galaxies with visible interaction signatures. Both NA10 categories have sub-levels: paired objects are sorted by relative separation (close, projected, apparent, or overlapping pairs), and interactions by morphology (disturbed, warp, shells, tails, or bridges).

In NA10, 22.3% of galaxies are paired with another object; of these, 72% are close pairs. Interacting galaxies are a much smaller subset, comprising only 7% of the NA10 sample. In GZ2, only 252 galaxies have $p_{odd} > 0.8$ and $p_{merger} > 0.8$. Using the raw number of votes, 3% of the NA10 paired galaxies have at least 10 GZ2 votes for a merger.

Figure 14 shows the distributions of NA10 paired and interacting galaxies that responded "yes" to Task 06. Most galaxies in NA10 and GZ2 have no votes for a merger, with only 6% of galaxies having $N_{merger} \geqslant 5$. Of the remaining galaxies, the numbers of both paired and interacting galaxies identified by NA10 begin to exceed the non-interacting population at a merger fraction above $p_{merger} > 0.25$. There is a significant population of non-interacting galaxies up to very high GZ2 vote fractions, however, which means that a simple cutoff is insufficient to produce a pure merger population by their criteria.

We visually examined galaxies that have high GZ2 merger fractions ($p_{merger} > 0.5$) but are classified by NA10 as non-interacting. The majority show obvious nearby companions, many of which appear to be tidally stripped or otherwise deformed. Some of these galaxies are likely the result of projection effects and are not truly interacting pairs – however, a significant fraction may be true interactions not accounted for in NA10. The contrary case (galaxies identified as interacting by NA10, but with no merger votes in GZ2), generally show faint extended features – mostly shells and tidal tails – that are clear signs of interacting. Most of these galaxies have no apparent companion visible in the image, however.

EFIGI does not have a dedicated category for mergers; galaxies are individually classified on whether they have any close companions ("contamination") or distortions in the galaxy profile ("perturbation"), which may or may not be merger-related. Galaxies cleanly classified by GZ2 as mergers are only weakly correlated with both attributes; the mean

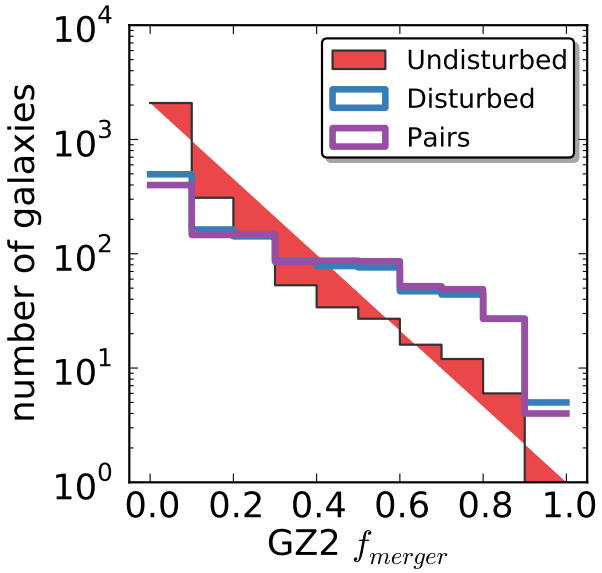

Figure 14. Merger classifications in GZ2 and NA10. Data are for the 3,878 galaxies in both samples with $p_{odd} > 0.223$, showing the distribution of the vote fraction for the "merger" response to Task 08 in GZ2. The majority of all galaxies have $p_{merger} < 0.1$. Galaxies classified by NA10 both as disturbed and in pairs dominate at $p_{merger} > 0.5$, but there remains a significant population of undisturbed galaxies even at the highest vote fractions.

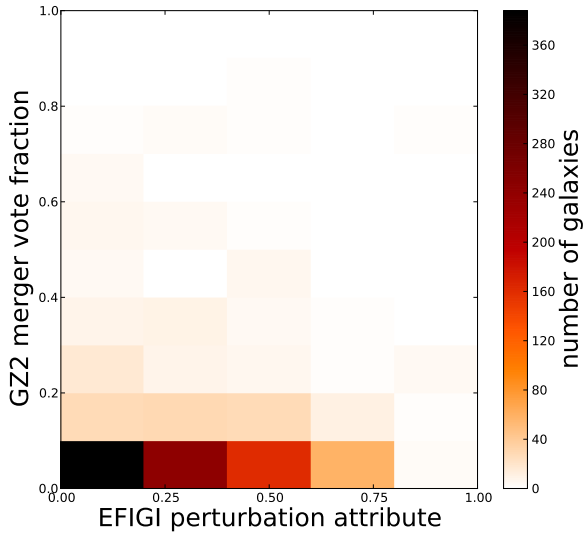

Figure 15. EFIGI perturbation classifications compared to GZ2 merger vote fractions. Data are for the 1,080 galaxies in both EFIGI and GZ2 with at least 10 responses to Task 08 (odd feature).

EFIGI value in GZ2 mergers is 0.31 for perturbation and 0.48 for contamination. Figure 15 shows that while there is only a very weak correlation ($\rho = 0.14$) between EFIGI perturbation and GZ2 merger vote fraction. The highly-perturbed galaxies with low GZ2 merger vote fractions are mostly dwarf peculiar and irregular galaxies with no sign of tidal features or an interacting companion.

Results from both expert catalogs are consistent with Casteels et al. (2013), who found that the mean vote fraction for mergers increases with decreasing projected separations (r_p), but then drops off significantly for the closest pairs at $r_p < 10$ kpc. At these separation, the GZ2 votes for Task 08 go instead to the "irregular" and "disturbed" responses.

5.2.4 T-types

Figure 16 shows the percentage of galaxies identified as having either a disk or features from the first question in the GZ2 tree, colour-coded by their NA10 T-types. There is a clear separation in the GZ2 fractions for galaxies classified as E versus Sa–Sd. Disk galaxies, including S0, have a median fraction for the GZ2 "features or disk" question of 0.80, with a standard deviation of 0.29. This distribution is bimodal, with one peak near 0.95 and another at 0.10. Disks with few GZ2 votes for "feature" are found to be primarily lenticular (S0) galaxies. If only galaxies with T-types Sa or later are considered, the peak at lower GZ2 vote fractions disappears. The median GZ2 vote fraction for these galaxies is 0.88, with a standard deviation of 0.23. The highest GZ2 vote fraction for an elliptical galaxy in NA10 is 0.741; therefore, any cut above this includes galaxies *exclusively* identified by NA10

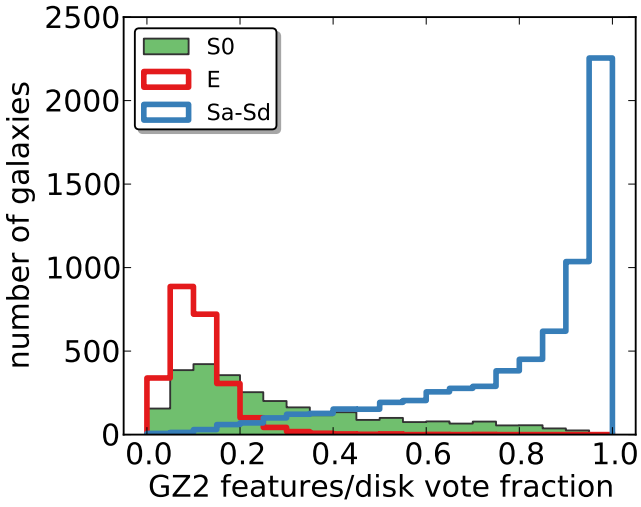

Figure 16. T-type classifications for NA10 and GZ2. Data are for the 12,480 galaxies found in both samples. The distribution of GZ2 vote fractions is separated by their T-type classification from NA10. Both elliptical and late-type spirals are strongly correlated with their GZ2 vote fraction. S0 galaxies are more commonly classified as ellipticals, but have a significant tail of high GZ2 features/disk vote fractions.

as late-type. Even if the confidence of this decreases for the larger GZ2 sample due to the inclusion of fainter galaxies, the previous limit of 0.8 (which may be conservative) reproduces the broad morphological cuts of NA10.

Since few objects are identified as stars or artifacts in the first GZ2 question, the vote fraction for smooth galaxies is approximately $p_{smooth} = (1 - p_{features/disk})$. Elliptical galaxies have a median vote fraction for the GZ2 "smooth"

question of 0.86 ± 0.07. The GZ2 votes for the NA10 ellipticals are more sharply peaked than NA10 late-types, lacking the long tail seen even for the very late types. This means that a cut on GZ2 votes for smooth galaxies at 0.8, for example, would include 4% late-type galaxies (20% if S0 galaxies are included).

For galaxies identified as NEDs, GZ2 users then vote on whether the galaxy has visible spiral structure (Task 04). For the few NA10 elliptical galaxies that have votes for this question, 85% have GZ2 vote fractions of zero, with the remainder weakly clustered around $p_{spiral} \sim 0.3$. For NA10 late-type galaxies, the majority of disk/feature objects have high GZ2 spiral structure vote fractions. For galaxies with at least 10 votes on Task 04, 70% of Sa or later-types have $p_{spiral} > 0.8$ from GZ2. This drops to 60% if S0 galaxies are included as late-type. The missing population is thus made up of galaxies that NA10 classify as having significant spiral structure, but for which GZ2 users cannot distinguish spiral arms. One might expect these galaxies to have lower magnitudes or surface brightnesses compared to the rest of the sample, thus lowering the confidence of GZ2 votes (there is no analog parameter associated with NA10 classifications). However, the apparent g and r magnitudes, as well as the absolute g-band magnitude, show no difference between galaxies above and below the 80% cutoff. Changing the value for the GZ2 vote fraction does not affect the results, so it appears that lower GZ2 vote fractions for spirals indicate intrinsically weaker (or less clearly-defined) spiral arms.

For disk galaxies with spiral structure, Task 10 in GZ2 asked users to classify the "tightness" of the arms. This had three options: tight, medium, or loose (accompanied with icons illustrating example pitch angles; see Figure 1). This allows investigation of the parameters which contribute to the Hubble classification of late-type galaxies which depends on both spiral arm and bulge morphology; tight spirals are presumed to be Sa/Sb, medium spirals Sb/Sc, and loose spirals Sc/Sd. The agreement between the GZ2 classification can be compared to Hubble types by using the NA10 classifications.

The left side of Figure 17 shows the distribution of NA10 T-types for galaxies based on their GZ2 vote fractions for winding arms. Vote fractions for both tight and medium winding arms are relatively normally distributed, with tight spirals peaking near 0.46 and medium spirals at 0.37. Strongly-classified loose spirals are much rarer, with 75% of galaxies having a vote fraction of less than 0.2. Almost no elliptical galaxies from the NA10 catalogue are included, although there are significant numbers of S0 galaxies.

For tight spirals, the category of galaxies with the highest vote fractions has more earlier-type spirals than galaxies with a low vote for tight spiral winding arms. For a tight spiral vote fraction above 0.9, 85% of galaxies are Sb or earlier. Medium-wound spirals with high vote fractions tend to be Sb and Sc galaxies – the proportion of both types increases as a function of medium-wound vote fraction, and constitute 84% of galaxies when $p_{medium} > 0.6$. Galaxies strongly classified as medium-wound are rare, however, with only 23 galaxies having $p_{medium} > 0.8$. Loose spirals are dominated by Sc and Sd galaxies at high vote fractions, comprising more than 50% of galaxies with $p_{loose} > 0.7$. Casteels et al. (2013) found that galaxies with $p_{loose} > 0.6$ often show tidal features and host a significant proportion of interacting galaxies. This distribution may reflect the experimental design of GZ2, with volunteers preferring extreme ends of a distribution rather than an indistinct 'central' option.

EFIGI T-types (Figure 17) show similar trends with respect to the GZ2 arms winding data. Late-type spirals (Sc–Sd) constitute about half of disk galaxies with $p_{loose} > 0.5$, with early-type spirals (Sa–Sb) occupying a similar distribution at $p_{tight} > 0.5$. S0 galaxies show nearly a flat distribution of GZ2 spiral tightness vote fractions; this is unsurprising, since by definition there should be no pitch angle without the presence of spiral arms.

Overall, a clear trend is demonstrated for looser GZ2 spiral arms to correspond with later spiral T-types from expert classifications. High vote fractions are mostly Sa/Sb galaxies for tight winding, Sb/Sc galaxies for medium winding, and Sc/Sd galaxies for loose winding. Individual galaxies, however, can show significant scatter in their GZ2 vote fractions and do not always separate the morphologies on the level of the Hubble T-types.

Having considered the effect of spiral arm tightness, we examine the relationship between bulge morphology and T-type. Disk galaxies in GZ2 are also classified by the visible level of bulge dominance (Task 05), irrespective of whether spiral structure is also identified. This task has four options: "no bulge", "just noticeable", "obvious", and "dominant" (accompanied with pictograms that illustrated bulge sizes compared to face-on spiral arms).

The left side of Figure 18 shows the distribution of NA10 T-types for galaxies based on their GZ2 vote fractions for winding arms. This figure shows only galaxies with at least 10 votes on bulge prominence. Vote fractions for both the "no bulge" and "dominant" responses peak strongly near zero and tail off as the vote fraction increases. Responses to the middle options, "just noticeable" and "obvious", resemble normal distributions peaking near 0.5.

"No bulge" galaxies in GZ2 are dominated by Sc and Sd spirals for non-zero vote fractions. For vote fractions above 0.1, 81% of galaxies are Sc or later; this rises to 100% for vote fractions higher than 0.6. "Just noticeable" galaxies show a smooth change in T-type distribution; low vote fractions are dominated by S0 and Sa galaxies, while high vote fractions are Sb–Sd. "Obvious" bulge galaxies are almost a mirror image of the "just noticeable" votes; low vote fractions are Sb–Sd galaxies, and high vote fractions are S0–Sa galaxies. Among galaxies classified as "dominant", less than 10 galaxies have vote fractions above 0.6 (which are a diverse mix of S0, Sa, and Sd). Most remaining galaxies have dominant vote fractions of less than 0.1; the T-types of the remaining galaxies between 0.1 and 0.6 mostly contain S0 and Sa spirals.

The link to T-type is more sharply defined for bulge prominence than for spiral tightness, according to the NA10 classifications. Very clean samples of late-type (Sb–Sd) spirals can be selected using only the "no bulge" parameter; additional samples with $\sim$ 10% contamination can be selected with the "just noticeable" and "obvious" distributions. Early-type spirals and lenticulars at the same purity level can also be selected. Elliptical galaxies from NA10 that have bulge prominence classified in GZ2 are most often "dominant", but there is no obvious separation of ellipticals from disk galaxies based on this task alone.

EFIGI T-types also correlate strongly with the GZ2

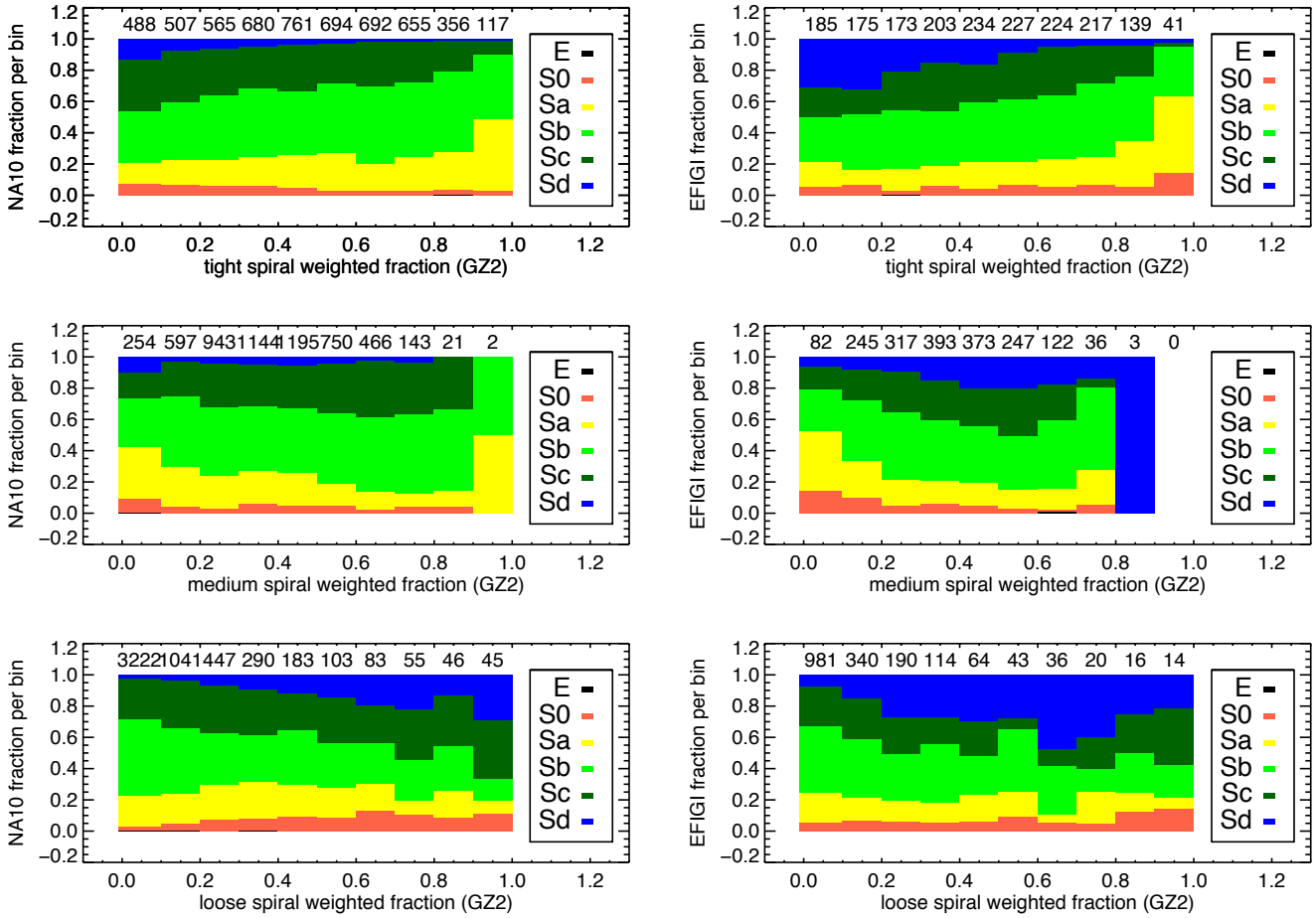

Figure 17. T-type classifications compared to the GZ2 vote fractions for spiral tightness (Task 10). Left side is NA10 T-types; right side is EFIGI T-types. Data are for the 5,515 (NA10) and 1,907 (EFIGI) galaxies, respectively, with at least 10 GZ2 votes for Task 10. The number of galaxies per bin is given along the top of each panel.

bulge dominance vote fractions. More than 90% of galaxies with $p_{nobulge} > 0.5$ are late-type spirals, with the bulk of these Sd galaxies. Both $p_{justnoticeable}$ and $p_{obvious}$ show a continuum of T-types as the vote fractions increase, with Sc and Sd galaxies having high vote fractions for the former and S0, Sa, and Sb galaxies in the latter. Galaxies with high vote fractions for $p_{dominant}$ are primarily S0s, along with a few elliptical galaxies that had enough votes as disk galaxies in GZ2 to answer the bulge classification question.

The agreement between NA10 and EFIGI T-types has been described by Baillard et al. (2011); both classification schemes primarily follow the same conventions of the RC3 system. EFIGI has slightly higher fractions of early-type spirals in less bulge-dominated galaxies than NA10, but it is unclear whether this is a result of different sample selection or individual bias on the part of the respective classifiers.

Since Hubble types are based on both the relative size of the bulge and the extent to which arms are unwound (Hubble 1926), we explored whether the combination of Tasks 05 and 10 from GZ2 can be mapped directly to T-types. We attempted to fit the numerical T-types to a linear combination of the GZ2 vote fractions for the bulge dominance and arms winding tasks. The best-fit result using symbolic regression (Schmidt & Lipson 2009), however, depends *only* on parameters relating to bulge dominance:

$$\text{T-type} = 4.63 + 4.17 \times p_{nobulge} - 2.27 \times p_{obvious} - 8.38 \times p_{dominant}. \quad (19)$$

Note that the $p_{justnoticeable}$ is implicitly included in this equation since the vote fractions for Task 05 must sum to 1. Inclusion of any vote fractions for arms winding responses made no significant difference in the r^2 goodness-of-fit metric.

This technique assumes that the difference in morphology is well-defined by mapping T-types to a linear scale, which is far from being justified. Figure 19 shows the distribution of the GZ2-derived T-type from Equation 19 compared to the NA10 values. The large amounts of overlap between adjoining T-types show that this clearly does not serve as a clean discriminator. S0 galaxies in particular would mistakenly be judged as significantly later types (Sa) on average using only this metric. One could make a cut between the earliest (Sa) and latest (Sd) spiral types based only on the vote fractions. Alternatively, the relative numbers galaxies could be used as the weights to construct the *probability* of a given T-type. This has yet to be conclusively tested.

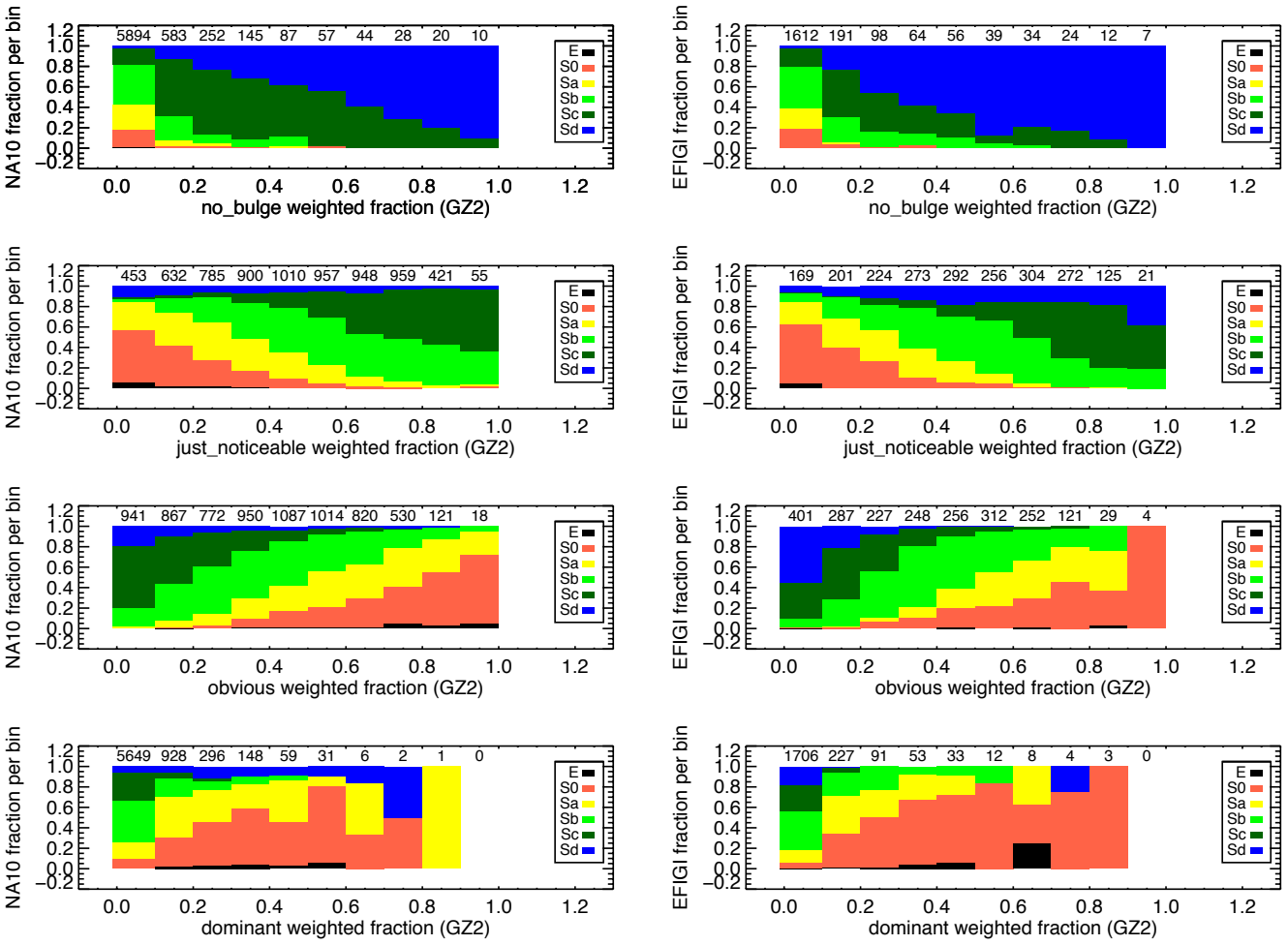

Figure 18. T-type classifications compared to the GZ2 vote fractions for bulge prominence (Task 05). Left side is NA10 T-types; right side is EFIGI T-types. Data are for the 7,120 (NA10) and 2,321 (EFIGI) galaxies, respectively, with at least 10 GZ2 votes for Task 05. The number of galaxies per bin is given along the top of each panel.

Finally, we note that Simmons et al. (2013) identified a significant effect in which nuclear point sources, such as AGN, can mimic bulges in the GZ2 classifications. This has not yet been accounted for in this analysis, but could potentially be addressed by separating the sample into AGN and quiescent galaxies (via BPT line ratios) and looking for systematic differences between the two samples.

5.2.5 Bulge prominence

EFIGI measures the bulge/total light ratio (B/T) in each galaxy, with the attribute strength corresponding to the relative contribution of the bulge. Elliptical galaxies have $B/T = 1$ and irregular galaxies $B/T = 0$. Baillard et al. (2011) show that B/T is correlated with arm curvature and anti-correlations with the presence of flocculent structure and hot spots, consistent with moving along the Hubble sequence.

Figure 20 (left panels) show the relationship between B/T and the four bulge dominance vote fractions from GZ2 for NED galaxies. $p_{obvious}$ is strongly correlated ($\rho = 0.65$) with B/T, while $p_{just\ noticeable}$ has a corresponding anti-correlation. Very few galaxies in the sample have either $p_{no\ bulge} > 0$ or $p_{dominant} > 0$, but those that do show corresponding changes in the EFIGI B/T. In particular, the number of galaxies with $B/T = 0$ and $p_{just_noticeable} > 0$ reinforces the results of Simmons et al. (2013), who showed that GZ2 bulge prominences increase with the presence of central point sources in the image (such as AGN).

5.2.6 Arm curvature

EFIGI also measures the arm curvature of each galaxy, with classifications very similar to the "tightness of spiral arms" question (Task 10) in GZ2. If both tasks and classifiers agree, one would expect galaxies with high GZ2 vote fractions for tight spirals to have EFIGI classifications at 0.75–1.0; GZ2 galaxies classified as medium spirals to be centered around 0.5; and loose spirals to have arm curvatures of 0.0–0.25.

The EFIGI arm curvature classifications broadly follow the trends expected from matching targets with GZ2. p_{tight} is the most strongly correlated with the EFIGI arm curvature parameter (Figure 20, right panels). The Spearman's correlation coefficient for tight spirals is $\rho = 0.62$. The medium spiral vote fraction is clustered in the middle of the EFIGI values, where galaxies with the highest GZ2 vote

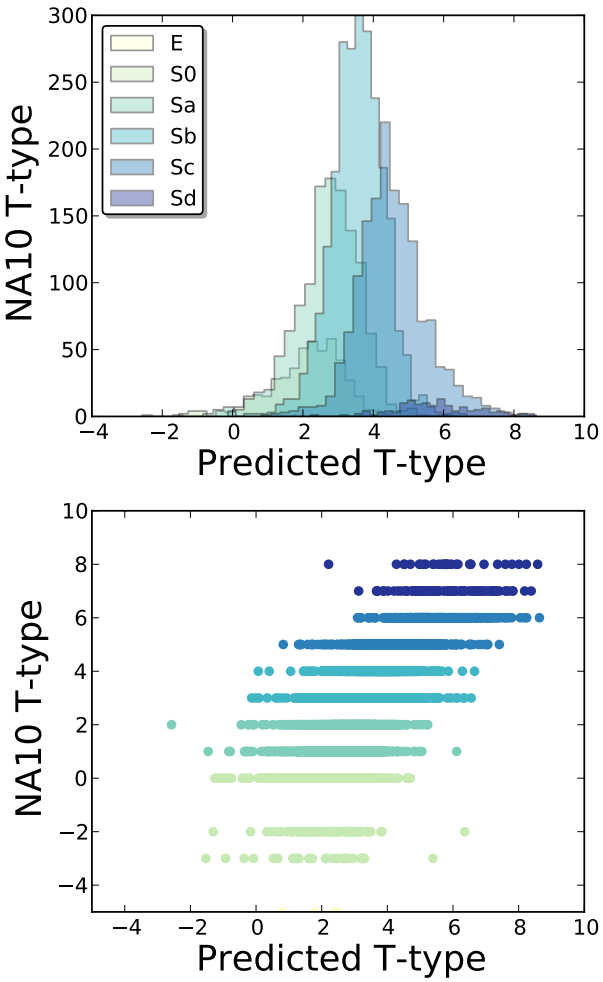

Figure 19. Predicted T-type classifications as fit by symbolic regression to the GZ2 data. Galaxies are colour-coded by their morphologies as identified by NA10. The top panel shows the histogram of predicted T-type based on Equation 19. The bottom shows the predicted T-types plotted against their NA10 values. Galaxies shown are only those with sufficient answers to characterize the arms winding and arms number GZ2 tasks, which selects heavily for late-type galaxies. This explains the lack of ellipticals in the plot, but highlights the fact that S0 galaxies do not agree well with the linear sequence. This is consistent with the "parallel-sequence" model of van den Bergh (1976) and revised by Kormendy & Bender (2012).

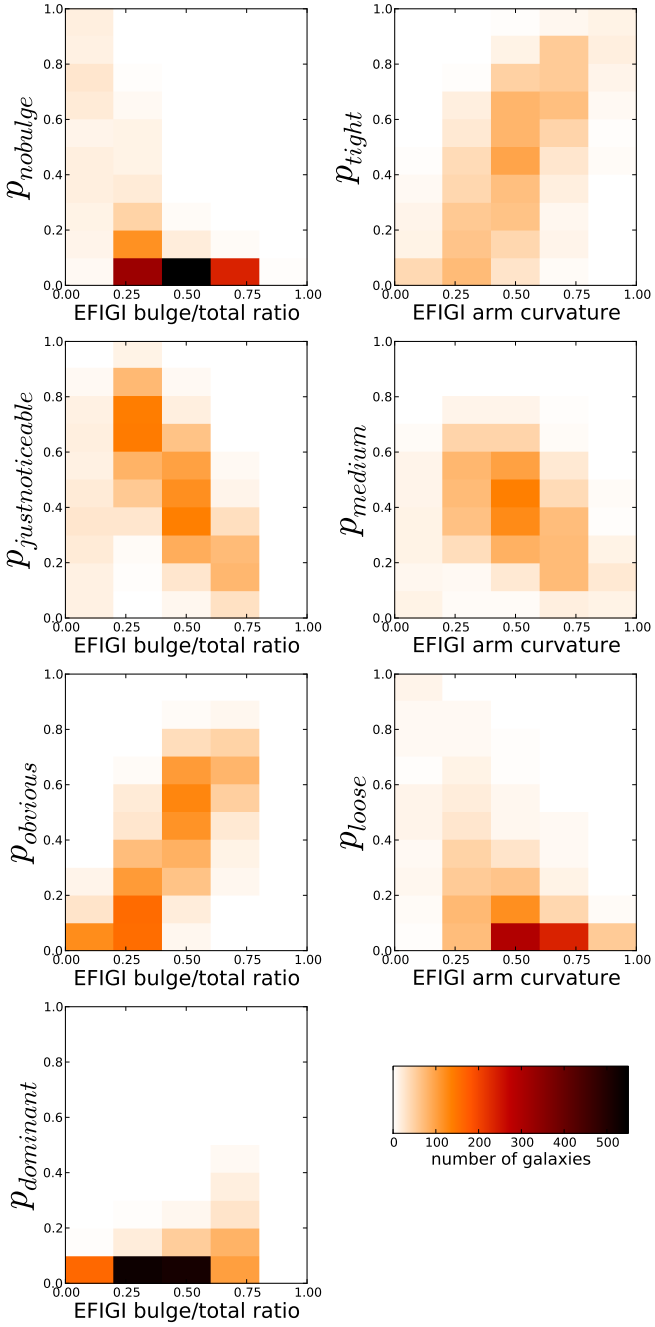

Figure 20. Left: EFIGI bulge/total ratio attributes compared to GZ2 vote fractions for "bulge prominence". Right: EFIGI arm curvature attributes compared to GZ2 vote fractions for "arms winding". Data are for the 1,544 NED galaxies in both samples.

fraction have EFIGI values of 0.25–0.50, with $\rho = -0.26$. Loose spirals shows an anti-correlation ($\rho = -0.54$); very few galaxies have GZ2 vote fractions above 0.5, but those which do have low EFIGI arm curvature values (0.0–0.25).

5.3 Automated classifications

Huertas-Company et al. (2011, HC11) have generated a large set of morphological classifications for the SDSS spectroscopic sample using a Bayesian automated routine. The broad nature of their probabilities (four broad morphological categories), do not directly relate to the majority of the GZ2 fine structure questions, such as bar or spiral arm structure. Comparison between the two samples, however, is useful to demonstrate the effect that smaller-scale features may have on automatically-classified morphologies.

The sample classified by HC11 is limited to galaxies with $z < 0.25$ that have both good photometric data and clean spectra. Their total of 698,420 galaxies is approximately twice the size of GZ2. The HC11 sample goes to fainter magnitudes, with more than 400,000 galaxies below the GZ2 limit of $m_r > 17$. Their morphological classification algorithm is implemented with support vector machine (SVM) software that tries to find boundaries between points

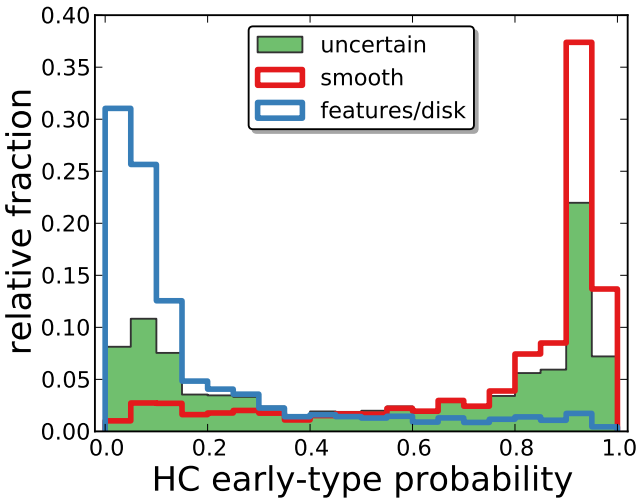

Figure 21. Distribution of HC11 early-type probabilities for galaxies split by their GZ2 classification. Data for smooth and features/disk are for galaxies with "clean" flags in Table 5; the uncertain classifications comprise galaxies with no flags set for Task 01.

in N-dimensional space, where N is determined by criteria including morphology, luminosity, colour, and redshift (Huertas-Company et al. 2008). The training set is the 2,253 galaxies in Fukugita et al. (2007), which are already classified by T-type. Each galaxy is assigned a probability of being in one of four subclasses: E, S0, Sab, and Scd (the latter two combining their respective late-type categories).

The inclusion of colour means that the HC11 classifications are not purely morphological, but rather include information about present-day star formation as well as the dynamical history which determines morphology. Studies of red spiral (Masters et al. 2010) and blue elliptical galaxies (Schawinski et al. 2009), for example, demonstrate the advantages of keeping these criteria separate.

Huertas-Company et al. (2011) directly compared their results to the GZ1 sample from Lintott et al. (2011). They found that robust classifications in GZ1 (flagged in our clean sample as being either confirmed ellipticals or spirals) have median probabilities of 0.92 according to their algorithm, indicating that sure GZ1 classifications are also sure in their catalogue. They also showed a near-linear relationship between the GZ1 debiased vote fraction and the HC11 probabilities. This is one of the first independent confirmations that the vote fractions may be related to the actual *probability* of a galaxy displaying a morphological feature.

Since the GZ2 galaxies are a subset of the GZ1 sample, the results for Task 01 are expected to be similar to those described in Huertas-Company et al. (2011). Figure 21 shows the distributions of the HC11 early- and late-type probabilities for GZ2 galaxies robustly identified ($f > 0.8$) as either smooth or having features/disks. The median HC11 early-type probability for GZ2 ellipticals is 0.85, and the late-type probability for GZ2 spirals is 0.95. This confirms the result that robust classifications in Galaxy Zoo agree with the automated algorithm.

An exception to this is a population of galaxies classified as "smooth" by GZ2 users, but which have very low early-type probabilities from HC11 (the bump on the left side of the first panel in Figure 21). The mean GZ2 vote fraction for these galaxies is consistent with those with high early-type probabilities – these galaxies are not marginally classified as ellipticals in GZ2. The roundness of the galaxy (Task 07 in GZ2) seems to play some role, as the low-HC11 smooth galaxies have fewer round galaxies and many more "cigar-shaped" galaxies in this sample. A high axial ratio might train the HC11 algorithm to infer the existence of a disk; the absence of any obvious spiral features or bulge/disk separation (verified by eye in a small subsample of the images) lead GZ2 users to categorise them as "smooth". There is a clear dependence on apparent magnitude; the lower peak disappears if only galaxies with $r < 16$ are plotted. The lower peak is also significantly bluer than the higher peak, with respective colours of $(g-r) = 0.67$ and $(g-r) = 0.97$. Since the SVM method does include SDSS colours as a parameter, we conjecture that the low HC11 early-type probability is in part due to the fact that they are blue, in addition to morphological features such as shape and concentration.

The right panel of Figure 21 shows the distribution of "unclassified" galaxies, for which none of the responses for Task 01 had a vote fraction > 0.8. The HC11 probability for these galaxies is bimodal, with the larger fraction classified as HC11 late-type and a smaller fraction as HC11 early-type.

Similar to the results from NA10 and EFIGI, the morphology classifications in HC11 have a strong dependence on bulge dominance (as measured from GZ2). Figure 22 shows the HC11 late-type spiral probability for disk galaxies as a function of vote fraction for the four GZ2 categories of bulge dominance. Since the majority of galaxies have both low $p_{nobulge}$ and $p_{dominant}$, the automated probabilities are primarily flat. There is a slight correlation between no bulge and later-type galaxies – even at $p_{nobulge} = 0.8$, though, the HC11 algorithm gives galaxies roughly equivalent probabilities between 0.2 and 0.8.

The relationship between bulge dominance and late-type probability is much stronger for the two intermediate responses for GZ2. Galaxies for which $p_{justnoticeable} > 0.6$ have a rapid increase in their late-type probabilities, with a sharp transition from the constant late-type probability between 0.25 and 0.6. As expected, the opposite effect occurs for obvious bulges; a vote fraction of $p_{obvious} < 0.2$ gives a very strong probability of being an Scd galaxy, while galaxies with $p_{obvious} > 0.5$ are favored to be classified as Sb or earlier.

Finally, we examined the potential effect of bars on the automated classifications. Figure 23 shows the average HC11 probability as a function of GZ2 bar vote fraction for NED galaxies. The relative proportions of galaxies as classified by HC11 is flat as a function of GZ2 p_{bar}, at 31% for early-type and 69% late-type. The presence of a bar thus does not strongly affect automated classifications, at least on the broad levels of early- and late-type.

6 CONCLUSIONS

We present the data release for the Galaxy Zoo 2 (GZ2) project. GZ2 uses crowd-sourced votes from citizen scientist classifiers to characterize morphology of more than 300,000 galaxies from the SDSS DR7. The images being classified are

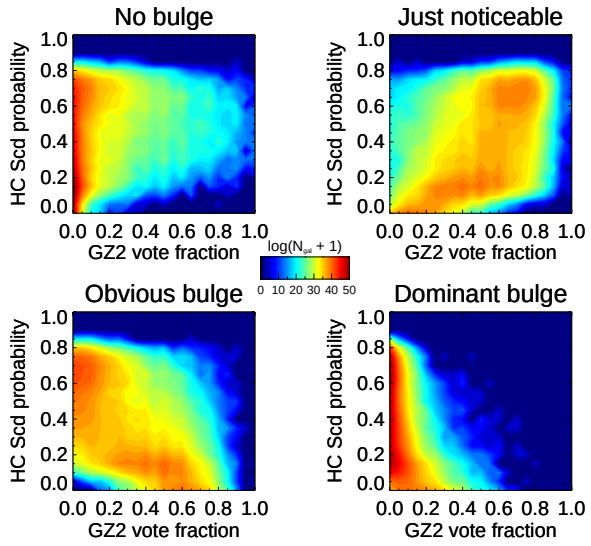

Figure 22. Huertas-Company et al. (2011) late-type spiral probability as a function of the GZ2 vote fraction for bulge dominance. The colour of the contours is log $(N_{gal} + 1)$, where N_{gal} ranges from 0 to 1.5×10^3. Galaxies shown are the 54,987 NED galaxies appearing in both samples.

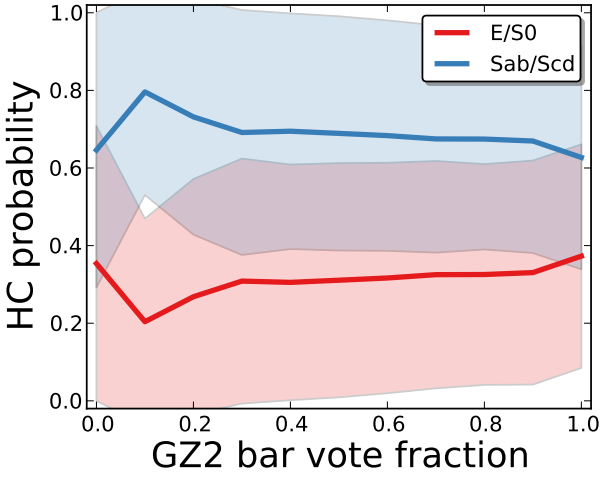

Figure 23. HC11 probabilities as a function of GZ2 bar vote fraction for 54,987 NED galaxies. Points give the mean probability in each bin of 0.1 width; shaded areas give the measured 1σ standard deviation.

gri colour composites, with the majority of galaxies selected on the basis of magnitude ($m_r < 17$), angular size ($r_{90} > 3''$), and redshift ($0.0005 < z < 0.25$) cuts. Deeper images from Stripe 82 are also included at both normal and coadded image depths.

GZ2 expands on the original Galaxy Zoo results by classifying a large array of fine morphological structures. In addition to previous distinctions between elliptical and spiral galaxies, GZ2 identifies the presence of bars, spiral structure, dust lanes, mergers, disturbed/interacting morphologies, and lenses. It also quantifies the relative strengths of galactic bulges (both edge-on and face-on), the tightness and multiplicity of spiral arms, and the relative roundness of elliptical galaxies. Classification was done via a multi-step decision tree presented to users in a web-based interface. The final catalogue is the result of nearly 60 million individual classifications of images.

Data reduction for the catalog begins by adjusting the effects of individual classifiers. Repeat classifications of objects by the same user are omitted from the catalog, and then an iterative weighting scheme is applied to users for each task based on their overall consistency. We then combine votes for each galaxy to assemble the overall classification; the strength of a particular feature is typically measured by the fraction of votes for a particular response (among all possible responses). The nature of the GZ2 classification scheme means that these vote fractions akin to conditional probabilities, however – for example, a galaxy must first be identified both as possessing a disk *and* as being "not edge-on" to measure the bar vote fraction.

Vote fractions for each response are also adjusted for classification bias, the effect of fine morphological features being more difficult to detect in smaller and fainter galaxies. Corrections to determine the debiased vote fractions are derived directly from the data itself.

The final catalog consists of five tables, comprising morphological classifications for the GZ2 main sample (separated into galaxies with spectroscopic and photometric redshifts) and galaxies from Stripe 82 (for normal-depth and two sets of coadded images). Data for each galaxy includes both the raw and weighted number of votes for each response, the raw and weighted vote fractions for each response, the debiased vote fraction, and an optional flag which indicates if a feature meets sufficient thresholds to be robustly identified. Portions of the data are presented in Tables 5–9; full machine-readable tables are available at `http://data.galaxyzoo.org` and in CasJobs following the SDSS Data Release 10.

We have carried out detailed comparisons of the GZ2 classifications to several other morphological catalogs. Early and late-type classifications are consistent with results from the original Galaxy Zoo, especially for galaxies in the clean samples. Overlap with expert catalogs (Nair & Abraham 2010; Baillard et al. 2011) shows good agreement for galaxies with medium to strong bars; GZ2 is less confident in identifying expert-classified weak and/or nuclear bars. Pairs and interacting galaxies are more difficult to reliably cross-match in a clean sample, although Casteels et al. (2013) have shown that the GZ2 "loose winding arms" parameter is a reliable proxy for interaction. Using expert classification of T-types for the overlapping subsamples, we show that the GZ2 bulge dominance parameter correlates strongly with Hubble type. Adding GZ2 measurements of the spiral arm tightness (corresponding to pitch angle) does not result in any significant increase in accuracy. Automated classifications from Huertas-Company et al. (2011) agree well with GZ2 in separating elliptical and late-type spirals, although S0 galaxies have significant overlaps between the two categories.

GZ2 contains more than an order of magnitude more galaxies than the largest comparable expert-classified catalogues (NA10, EFIGI) while still classifying detailed morphological features not replicable by automated classifications. GZ2 data have already been used to demonstrate a relationship between bar fraction and the colour, gas fractions, and bulge size of disk galaxies (Masters et al. 2011, 2012), as well as studies of the bar color and length itself

Hoyle et al. (2011). The size of the catalogs has allowed for the discovery and study of comparatively rare objects, such as early-type dust lane galaxies (Kaviraj et al. 2012) and bulgeless AGN hosts (Simmons et al. 2013). Direct use of the GZ2 likelihoods has also been used to quantify the environmental dependence on morphology, showing a correlation for barred and bulge-dominated galaxies (Skibba et al. 2012) and identifying reliable signatures of interaction from GZ2 data (Casteels et al. 2013).

The success of the Galaxy Zoo project has already shown that volunteer classification is a robust method for analysis of large data sets; this public release of the detailed Galaxy Zoo 2 catalogue intends to build on this further, both from their inherent scientific value and as training sets for refining methods of automatic classification techniques.

ACKNOWLEDGMENTS

The data in this paper are the result of the efforts of the Galaxy Zoo 2 volunteers, without whom none of this work would be possible. Their efforts are individually acknowledged at `http://authors.galaxyzoo.org`.

The development of Galaxy Zoo 2 was supported by The Leverhulme Trust. CJL acknowledges support from the STFC Science in Society program. KS gratefully acknowledges support from Swiss National Science Foundation Grant PP00P2_138979/1.

Boilerplate for individual grants, Zooniverse support, etc.

STFC numbers for everyone in the UK.

This research made use of Montage, funded by the National Aeronautics and Space Administration's Earth Science Technology Office, Computation Technologies Project, under Cooperative Agreement Number NCC5-626 between NASA and the California Institute of Technology. Montage is maintained by the NASA/IPAC Infrared Science Archive. It also made extensive use of the Tool for OPerations on Catalogues And Tables (TOPCAT), which can be found at `http://www.starlink.ac.uk/topcat/` (Taylor 2005, 2011).

Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the U.S. 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 website 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

Abazajian K. N. et al., 2009, ApJS, 182, 543
Aguerri J. A. L., Méndez-Abreu J., Corsini E. M., 2009, A&A, 495, 491
Annis J. et al., 2011, ArXiv e-prints
Baillard A. et al., 2011, A&A, 532, A74
Bamford S. P. et al., 2009, MNRAS, 393, 1324
Banerji M. et al., 2010, MNRAS, 406, 342
Barazza F. D., Jogee S., Marinova I., 2008, ApJ, 675, 1194
Blanton M. R. et al., 2003, ApJ, 592, 819
Buta R., Combes F., 1996, Fund. Cosm. Phys., 17, 95
Casteels K. R. V. et al., 2013, MNRAS, 429, 1051
Csabai I. et al., 2003, AJ, 125, 580
Darg D. W. et al., 2010, MNRAS, 401, 1043
Davis D., Hayes W., 2013, ArXiv e-prints
de Vaucouleurs G., de Vaucouleurs A., Corwin, Jr. H. G., Buta R. J., Paturel G., Fouqué P., 1991, Third Reference Catalogue of Bright Galaxies. Springer-Verlag
Fischer D. A. et al., 2012, MNRAS, 419, 2900
Fukugita M. et al., 2007, AJ, 134, 579
Hinshaw G. et al., 2012, ArXiv e-prints
Hoyle B. et al., 2011, MNRAS, 415, 3627
Hubble E. P., 1926, ApJ, 64, 321
Hubble E. P., 1936, Realm of the Nebulae. Yale University Press
Huertas-Company M., Aguerri J. A. L., Bernardi M., Mei S., Sánchez Almeida J., 2011, A&A, 525, A157+
Huertas-Company M., Rouan D., Tasca L., Soucail G., Le Fèvre O., 2008, A&A, 478, 971
Jacob J. C. et al., 2010, Montage: An Astronomical Image Mosaicking Toolkit. Astrophysics Source Code Library
Kaviraj S. et al., 2012, MNRAS, 423, 49
Kormendy J., Bender R., 2012, ApJS, 198, 2
Kormendy J., Kennicutt, Jr. R. C., 2004, ARA&A, 42, 603
Lintott C. et al., 2011, MNRAS, 410, 166
Lintott C. J. et al., 2008, MNRAS, 389, 1179
Lupton R., Blanton M. R., Fekete G., Hogg D. W., O'Mullane W., Szalay A., Wherry N., 2004, PASP, 116, 133
Martig M., Bournaud F., Croton D. J., Dekel A., Teyssier R., 2012, ApJ, 756, 26
Masters K. L. et al., 2010, MNRAS, 405, 783
Masters K. L. et al., 2012, MNRAS, 424, 2180
Masters K. L. et al., 2011, MNRAS, 411, 2026
Nair P. B., Abraham R. G., 2010, ApJS, 186, 427
Nieto-Santisteban M. A., Szalay A. S., Gray J., 2004, in Astronomical Society of the Pacific Conference Series, Vol. 314, Astronomical Data Analysis Software and Systems (ADASS) XIII, Ochsenbein F., Allen M. G., Egret D., eds., p. 666
Sandage A., 1961, The Hubble atlas of galaxies. Carnegie Institute of Washington
Schawinski K. et al., 2009, MNRAS, 396, 818
Schmidt M., Lipson H., 2009, Science, 324, 81
Schwamb M. E. et al., 2012, ApJ, 754, 129
Simmons B. D. et al., 2013, MNRAS, 429, 2199
Simpson R. J. et al., 2012, MNRAS, 424, 2442

Skibba R. A. et al., 2012, MNRAS, 423, 1485
Smith A. M. et al., 2011, MNRAS, 412, 1309
Strauss M. A. et al., 2002, AJ, 124, 1810
Taylor M., 2011, TOPCAT: Tool for OPerations on Catalogues And Tables. Astrophysics Source Code Library
Taylor M. B., 2005, in Astronomical Society of the Pacific Conference Series, Vol. 347, Astronomical Data Analysis Software and Systems XIV, Shopbell P., Britton M., Ebert R., eds., p. 29
van den Bergh S., 1976, ApJ, 206, 883
York D. G. et al., 2000, AJ, 120, 1579

This paper has been typeset from a TEX/ LATEX file prepared by the author.

Table 5. Morphological classifications of GZ2 main sample galaxies with spectra

				t01_smooth_or_features_a01_smooth_						t01_smooth_or_features_a02_features_or_disk_						...
SDSS DR7 objID	sample	N_{class}	N_{votes}	count	wt_count	fraction	wt_fraction	debiased	flag	count	weight	fraction	wt_fraction	debiased	flag	
588017703996096547	original	44	349	1	0.1	0.023	0.002	0.002	0	42	42.0	0.955	0.975	0.975	1	
587738569780428805	original	45	185	5	5.0	0.111	0.115	0.115	0	38	38.0	0.844	0.873	0.873	1	
587735695913320507	original	46	372	0	0.0	0.000	0.000	0.000	0	44	44.0	0.957	0.966	0.966	1	
587742775634624545	original	45	289	8	8.0	0.178	0.178	0.178	0	37	37.0	0.822	0.822	0.822	1	
587732769983889439	extra	49	210	12	12.0	0.245	0.249	0.454	0	36	36.0	0.735	0.748	0.749	0	
588017725475782665	extra	42	149	27	27.0	0.643	0.686	0.771	0	12	12.0	0.286	0.305	0.305	0	
588017702391578633	original	45	356	0	0.0	0.000	0.000	0.000	0	45	45.0	1.000	1.000	1.000	1	
588297864730181658	original	45	206	4	4.0	0.089	0.091	0.091	0	39	38.5	0.867	0.871	0.871	1	
588017704545812500	original	43	360	0	0.0	0.000	0.000	0.000	0	43	43.0	1.000	1.000	1.000	1	
588017566564155399	extra	43	244	6	6.0	0.140	0.143	0.143	0	35	35.0	0.814	0.833	0.833	1	

Note. — The full, machine-readable version of this table is available at http://data.galaxyzoo.org. A portion is shown here for guidance on form and content. The full table contains 252,750 rows (one for every galaxy in the sample), and 226 columns, with six variables for each of the 37 GZ2 morphology classifications.

Table 6. Morphological classifications of GZ2 main sample galaxies with photo-z

				t01_smooth_or_features_a01_smooth_						t01_smooth_or_features_a02_features_or_disk_						...
SDSS DR7 objid	sample	N_{class}	N_{votes}	count	wt_count	fraction	wt_fraction	debiased	flag	count	weight	fraction	wt_fraction	debiased	flag	
587722981736579107	original	43	181	27	27.0	0.628	0.648	0.648	0	14	14.0	0.326	0.336	0.336	0	
587722981741691055	original	44	133	40	40.0	0.909	0.909	0.909	1	1	1.0	0.023	0.023	0.023	0	
587722981745819655	original	46	221	17	17.0	0.370	0.378	0.378	0	22	22.0	0.478	0.489	0.489	0	
587722981746082020	original	44	172	31	31.0	0.705	0.771	0.386	0	6	6.0	0.136	0.149	0.814	1	
587722981746344092	original	43	358	0	0.0	0.000	0.000	0.000	0	43	43.0	1.000	1.000	1.000	1	
587722981747982511	original	45	156	37	37.0	0.822	0.850	0.578	0	3	3.0	0.067	0.069	0.289	0	
587722981748375814	original	52	198	44	44.0	0.846	0.846	0.674	0	8	8.0	0.154	0.154	0.204	0	
587722981748768914	original	46	350	3	3.0	0.065	0.065	0.117	0	43	43.0	0.935	0.935	0.935	1	
587722981748768984	original	42	140	37	36.2	0.881	0.900	0.699	0	4	4.0	0.095	0.100	0.128	0	
587722981749031027	original	50	158	46	45.8	0.920	0.932	0.710	0	2	1.4	0.040	0.028	0.107	0	

Note. — The full, machine-readable version of this table is available at http://data.galaxyzoo.org. A portion is shown here for guidance on form and content, which are identical to those in Table 5.

Table 7. GZ2 morphological classifications of normal-depth images of Stripe 82 galaxies

			t01_smooth_or_features_a01_smooth_						t01_smooth_or_features_a02_features_or_disk_						...
Stripe82 objID	N_{class}	N_{votes}	count	wt_count	fraction	wt_fraction	debiased	flag	count	weight	fraction	wt_fraction	debiased	flag	
587730845812064684	46	135	38	38.0	0.826	0.851	0.842	1	2	2.0	0.043	0.045	0.049	0	
587730845812065247	49	230	26	26.0	0.531	0.551	0.551	0	20	19.1	0.408	0.405	0.405	0	
587730845812196092	48	368	2	2.0	0.042	0.042	0.042	0	46	45.3	0.958	0.958	0.958	1	
587730845812196825	42	177	26	26.0	0.619	0.633	0.633	0	15	15.0	0.357	0.365	0.365	0	
587730845812524122	51	149	48	48.0	0.941	0.961	0.961	1	0	0.0	0.000	0.000	0.000	0	
587730845812654984	49	201	33	33.0	0.673	0.673	0.638	0	15	15.0	0.306	0.306	0.348	0	
587730845812655541	46	193	25	25.0	0.543	0.543	0.543	0	14	14.0	0.304	0.304	0.304	0	
587730845812720365	43	152	34	33.8	0.791	0.790	0.721	0	8	8.0	0.186	0.187	0.226	0	
587730845812720699	46	233	24	22.5	0.522	0.506	0.506	0	18	18.0	0.391	0.404	0.404	0	
587730845812851385	45	147	39	39.0	0.867	0.884	0.837	1	3	3.0	0.067	0.068	0.100	0	

Note. — The full, machine-readable version of this table is available at http://data.galaxyzoo.org. A portion is shown here for guidance on form and content, which are identical to those in Table 5. Classifications here are for normal-depth images from Stripe 82, which goes to a deeper magnitude limit ($m_r > 17.7$) galaxies in the main sample.

Table 8. GZ2 morphological classifications of coadded images (set 1) of Stripe 82 galaxies

			t01_smooth_or_features_a01_smooth_						t01_smooth_or_features_a02_features_or_disk_						...
Stripe82 objID	N_{class}	N_{votes}	count	wt_count	fraction	wt_fraction	debiased	flag	count	weight	fraction	wt_fraction	debiased	flag	
8647474690312307154	20	74	15	14.4	0.750	0.742	0.749	0	3	3.0	0.150	0.155	0.139	0	
8647474690312307877	17	54	13	13.0	0.765	0.765	0.765	0	3	3.0	0.176	0.176	0.176	0	
8647474690312308880	12	32	10	10.0	0.833	0.833	0.833	1	0	0.0	0.000	0.000	0.000	0	
8647474690312373464	22	75	18	18.0	0.818	0.829	0.829	1	2	1.7	0.091	0.079	0.079	0	
8647474690312438284	23	149	3	3.0	0.130	0.136	0.136	0	17	17.0	0.739	0.769	0.769	0	
8647474690312505086	15	58	11	11.0	0.733	0.748	0.748	0	4	3.7	0.267	0.252	0.252	0	
8647474690312832559	20	77	14	14.0	0.700	0.700	0.781	0	5	5.0	0.250	0.250	0.206	0	
8647474690312898532	14	68	9	9.0	0.643	0.643	0.643	0	5	5.0	0.357	0.357	0.357	0	
8647474690312962734	21	77	15	15.0	0.714	0.714	0.679	0	5	5.0	0.238	0.238	0.228	0	
8647474690312963665	12	43	11	11.0	0.917	0.917	0.917	1	1	1.0	0.083	0.083	0.083	0	

Note. — The full, machine-readable version of this table is available at http://data.galaxyzoo.org. A portion is shown here for guidance on form and content, which are identical to those in Table 5. Classifications here are for the coadded images (set 1; see §2.2) from Stripe 82, which goes to a deeper magnitude limit and has a better angular resolution than galaxies in the main sample. There is no colour desaturation for background sky pixels in this set of images.

Table 9. GZ2 morphological classifications of coadded images (set 2) of Stripe 82 galaxies

			t01_smooth_or_features_a01_smooth_						t01_smooth_or_features_a02_features_or_disk_						...
Stripe82 objID	N_{class}	N_{votes}	count	wt_count	fraction	wt_fraction	debiased	flag	count	weight	fraction	wt_fraction	debiased	flag	
8647474690312307154	16	72	10	10.0	0.625	0.625	0.629	0	5	5.0	0.312	0.312	0.259	0	
8647474690312307877	21	84	17	17.0	0.810	0.810	0.810	1	4	4.0	0.190	0.190	0.190	0	
8647474690312308318	23	88	18	18.0	0.783	0.783	0.722	0	4	4.0	0.174	0.174	0.208	0	
8647474690312308880	16	48	16	16.0	1.000	1.000	1.000	1	0	0.0	0.000	0.000	0.000	0	
8647474690312373464	23	89	17	17.0	0.739	0.739	0.739	0	4	4.0	0.174	0.174	0.174	0	
8647474690312438284	11	91	0	0.0	0.000	0.000	0.000	0	11	11.0	1.000	1.000	1.000	1	
8647474690312505086	12	65	4	3.4	0.333	0.295	0.295	0	8	8.0	0.667	0.705	0.705	0	
8647474690312832559	23	75	14	14.0	0.609	0.629	0.666	0	4	4.0	0.174	0.180	0.175	0	
8647474690312898532	26	129	12	12.0	0.462	0.462	0.492	0	14	14.0	0.538	0.538	0.449	0	
8647474690312962734	20	69	18	17.0	0.900	0.895	0.840	1	2	2.0	0.100	0.105	0.111	0	

Note. — The full, machine-readable version of this table is available at http://data.galaxyzoo.org. A portion is shown here for guidance on form and content, which are identical to those in Table 5. Classifications here are for the coadded images (set 2; see §2.2) from Stripe 82, which goes to a deeper magnitude limit and has a better angular resolution than galaxies in the main sample. Pixels in the sky background are colour desaturated in this set of images.