Mon. Not. R. Astron. Soc. 000, 1–28 (2012) Printed 21 February 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], Steven P. Bamford[3], Karen L. Masters[4], Brooke D. Simmons[2], Kevin Schawinski[5], Lucy Fortson[1], Robert J. Simpson[2], Ramin A. Skibba[6], Edward M. Edmondson[4], Arfon M. Smith[2,7]

[1] *University of Minnesota, USA*
[2] *University of Oxford, UK*
[3] *University of Nottingham, UK*
[4] *University of Portsmouth, UK*
[5] *ETH, Zürich, Switzerland*
[6] *University of California San Diego, USA*
[7] *Adler Planetarium, USA*

Accepted 1988 December 15. Received 1988 December 14; in original form 1988 October 11

ABSTRACT
Morphology is a powerful and unique probe for quantifying the dynamical history of a galaxy. 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 number of galaxies in large samples make this impractical for individual astronomers. Galaxy Zoo 2 (GZ2) is a citizen science project that provides morphological classifications of more than 300,000 galaxies drawn from the Sloan Digital Sky Survey. These include all galaxies in the DR7 Legacy survey down to $r > 17$, along with deeper classifications of galaxies in Stripe 82. The original Galaxy Zoo project primarily separated galaxies only into early- or late-types; GZ2 classifies finer morphological features. These include the presence of bars, bulges, edge-on disks, and merging galaxies, as well as quantifying the strength of multiplicity of features such as galactic bulges and spiral arms. This paper presents the full data release for the project, including measures of classification accuracy and user bias. We show that the majority of GZ2 classifications agree with those made by salaried astronomers, especially for T-types, strong bars, and arm curvature. Both raw and reduced data products are fully available and can be obtained in electronic format at http://data.galaxyzoo.org.

multiplicity
strength

need to
agree on
term
professionally
trained

Key words: galaxies

abstract = 300,000

1 INTRODUCTION

The Galaxy Zoo project (Lintott et al. 2008) was launched in 2007 to provide morphological classifications of nearly one million galaxies drawn from the Sloan Digital Sky Survey (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 to a manageable size only simple morphological distinctions were initially requested, essentially dividing systems into elliptical, spiral and merger. This paper presents data and results from that project's successor, Galaxy Zoo 2 (GZ2), which collected more sophisticated morphological classifications for more than 250,000 of the brightest SDSS galaxies.[1]

While the morphological distinction used in Galaxy Zoo 1 (GZ1) – that which divides spiral and elliptical systems – is the most fundamental, there is a long history of finer grained morphological classification. The first systematic approach to classification (Hubble 1936)

* E-mail: willett@physics.umn.edu

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

included a division between barred and unbarred spirals, creating the famous 'tuning fork', and further distinctions based on the shape of early-type systems or tightness of spiral arms. These finer distinctions are 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 (a review is given in Kormendy & Kennicutt (2004) and an updated picture by 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) and so on. Careful classification of morphological features is thus essential if the assembly and evolution of the galaxy population is to be understood.

are

are these effects correlated? I think that is an open question

fine-scale

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. The largest detailed professional classification effort to date was undertaken by Nair & Abraham (2010a), who provide classifications of $\sim$ 14000 systems. The present study includes an order of magnitude more systems, allowing for a more careful study of the relationships and interdependence of such small scale morphological features.

what is meant by system? all boxes for mergers

The use of proxies for morphology such as colour, concentration index, spectral features, surface brightness prole file structural features, spectral energy distribution or some combination of these is not an adequate substitute; each proxy has an unknown biased relation with the morphological features under study. The complexity of the relationship between these variables is, rather, the main reason for requiring such a large set of classifications, as one can only isolate significant numbers of red, barred, bulgeless spirals in the field (to give one example) with a sufficiently comprehensive starting set.

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) XXX ADD MORE REFS XXX, 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.

but if they use GZ1 as input, this is not consistent

These results have been made possible by the participation in the Galaxy Zoo project by hundreds of thousands of 'citizen scientists'. Since the original Galaxy Zoo demonstrated the utility of this method in producing both scientifically-useful catalogues and serendipitous discoveries (see Lintott et al. (2011) for a review of Galaxy Zoo 1 results), this method has been expanded beyond simple classifications to use cases which include exoplanet discovery (Fischer et al. 2011; Schwamb et al. 2012) and a census of bubbles associated with star formation in the MWG (Simpson et al. 2012) amongst many others.

2 PROJECT DESCRIPTION

2.1 Sample selection

Objects classified for Galaxy Zoo 2 included around 250,000 of the brightest resolved galaxies from the SDSS North Galactic Cap region. The goal was to exclude the most distant, faintest and smallest systems within which fine morphological features would not be resolved. The sample was restricted to the SDSS DR7 'Legacy' catalogue (Abazajian et al. 2009), and therefore excludes observations made by SDSS for other purposes, such as the SEGUE survey.

Three further cuts were applied to the DR7 Legacy sample. A Petrosian magnitude brighter than 17.0 in the r-band (after Galactic extinction correction was applied) is required, along with a r-band Petrosian radius greater than 3 arcsec. Galaxies which had a spectroscopic redshift in the DR7 catalogue outside the range $0.0005 < z < 0.25$ were also excluded; 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 are also excluded. Galaxy Zoo 2 was launched with the resulting sample of 245,609 images on 2009-02-16. spell out

An error in the original query meant that the initial Galaxy Zoo 2 sample ("original" sample) was missing some objects ("extra" sample) on launch, specifically those flagged as both BLENDED and CHILD. These objects, which are typically slightly brighter, larger and bluer than the general population, were added to the site on 2009-09-02. The rate at which images from the respective samples were shown to users was tuned so that the "extra" sample would catch-up to the same number of classifications as the "original" sample.

? at what rate? by when?

In addition to the sample from the Legacy survey, we later added images from Stripe 82, a section along the celestial equator in the Southern Galactic Cap. These are the only images for which we included multiple images of individual galaxies: single-depth and co-added. The 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 single-depth imaging (Annis et al. 2011). This coadded sample was added after launch, together with matching galaxies from the normal-depth Stripe 82 imaging ("stripe82" sample). These were mostly added on 2009-09-02 (but note that $\sim$ 7700 of the "stripe82_coadd" sample were not added until 2010-09-24).

The initial Stripe 82 coadd images were visually very different from the normal SDSS images. In order to rectify this, another set of Stripe 82 coadd images were produced and added into the site on 2009-11-04. In the tables these two sets of images are indicated by sample names "stripe82_coadd_1" and "stripe82_coadd_2".

Need paragraph from SB or EE on the details for making the coadded images.

For most of the duration of Galaxy Zoo 2, all the images were shown to classifiers in a random order. However, we ~~desired to have all galaxies classified at least~~ required a minimum number ~~of times~~ of classifications for all galaxies. Therefore, in the final period of Galaxy Zoo 2, accompanied by a competition with a running tally (dubbed the Zoonometer), objects with low numbers of classifications were preferentially shown ~~in an attempt to get them up to~~

Table 1. GZ2 sample properties

Sample	$N_{galaxies}$	$N_{class.}$ median	m_r depth [mag]
original	245,609	44	17.0
extra	28,174	41	17.0
Stripe 82 normal	21,522	45	17.77
Stripe 82 normal (mag-limited)	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 maglim)	283,971	44	17.0

to increase classifications to

a minimum of 40 ~~classifications~~ for the "original", "extra" and "stripe82" samples, and 20 for the "stripe82_coadd_2" sample. The "stripe82_coadd_1" sample was removed from the site at this time. ~~This effort had mixed success, since separate classifications of the Galaxy Wars project also contributed to the count.~~ The main sample galaxies finished with a median of 44 classifications, with 27% having fewer than 40; the "stripe82_coadd_2" galaxies had a median of 21 classifications and 26% of them had fewer than 20 (Table 1). ? why

can get rid of?

The primary sample for GZ2 analysis consists of the combined "original", "extra", and the Stripe 82 normal-depth images for which $r \leqslant 17.0$. We have verified that there are no significant differences in classifications between these samples that could have been caused, for example, by a time-dependent bias (since the samples were introduced on different dates). This is hereafter referred to as the GZ2 **main sample.** Data from both the Stripe 82 normal-depth images with $r > 17.0$ and the two sets of coadded images are included as separate data products.

2.2 Decision tree

Data for Galaxy Zoo 2 was collected via a web-based interface. Users of the interface needed to register with a username for classifications to be recorded, but were not required to complete any tutorials. They were then shown a *gri* colour composite image of a galaxy generated from the SDSS ImgCutout web service (Nieto-Santisteban et al. 2004), with the image randomly chosen from our sample database.

Need text from CL, EE, AS, or SB on image cutout creation; pixel scale, dynamic range, size/centering of images, etc.

Classification of the galaxies proceeds via a multi-step tree. Each classification begins with a slightly modified version of the original Galaxy Zoo task, with users identifying whether the galaxy is either "smooth", has "features or a disk", or is a "star or artifact" in the image. Subsequent classification questions 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 question will not be asked if they select either of the other two options.

The Galaxy Zoo 2 tree has 11 classification tasks with a total of 37 possible responses (Table 2). A classifier selects only one option for each task, after which they are immediately taken to the next step in the tree. Task 01 is the only question answered for all objects in the sample. Once

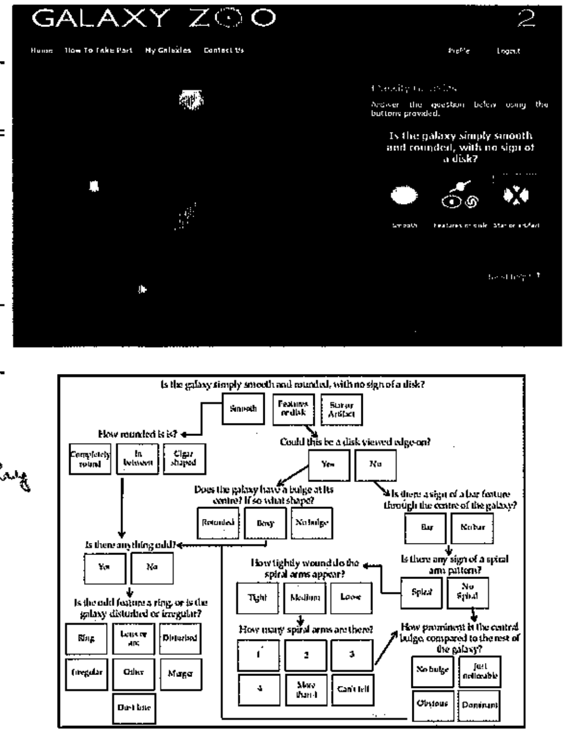

Figure 1. *Top:* Front page of the web interface for Galaxy Zoo 2, displaying Task 01. *Bottom:* Flowchart of the 11 classification tasks for GZ2, beginning at the top center.

a classification was completed, an image of the next galaxy is automatically displayed and the user can begin on a new object.

Data from the classifications was stored in a live Structured Query Language (SQL) database. In addition to the morphology classifications, we also registered the timestamp, user identification, and galaxy identification for each asset in the database.

The last GZ2 classifications were collected on 2010-04-29, spanning just over 14 months. The final dataset contained 16,340,298 classifications (comprising a total of 58,719,719 questions) by 83,943 participants.

with the project

3 DATA REDUCTION

3.1 Multiple classifications

In a small percentage of cases, an individual user may classify the same object more than once. Since we wish to treat each click as an independent measurement, we removed multiple classifications of the same object by a given user from the data, keeping only the last submitted classification. Such repeat classifications only occurred for ~~a small proportion of~~ ~1% of objects ~~(~ 1%)~~, and an even smaller proportion (~ 0.01%) significantly enough to potentially alter their classifications.

3.2 Consistency and invididual user weighting

The next step in reducing the data is to remove the influence of unreliable users. To do so we applied an iterainstead of "answers" should use "response" terminology throughout text

fix collision

Table 2. GZ2 classification tree

Task	Question	Responses	Next t
01	*"Is the galaxy simply smooth and rounded, with no sign of a disk?"*	smooth	→ 0
		features or disk	→ 0
		star or artifact	enc
02	*"Could this be a disk viewed edge-on?"*	yes	→ 0
		no	→ 0
03	*"Is there a sign of a bar feature through the center of the galaxy?"*	yes	→ 0
		no	→ 0
04	*"Is there any sign of a spiral arm pattern?"*	yes	→ 1
		no	→ 0
05	*"How prominent is the central bulge, compared with the rest of the galaxy?"*	no bulge	→ 0
		just noticeable	→ 0
		obvious	→ 0
		dominant	→ 0
06	*"Is there anything odd?"*	yes	→ 0
		no	enc
07	*"How rounded is it?"*	completely round	→ 0
		in between	→ 0
		cigar-shaped	→ 0
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	→ 10
11	*"How many spiral arms are there?"*	1	→ 05
		2	→ 05
		3	→ 05
		4	→ 05
		more than four	→ 05
		can't tell	→ 05

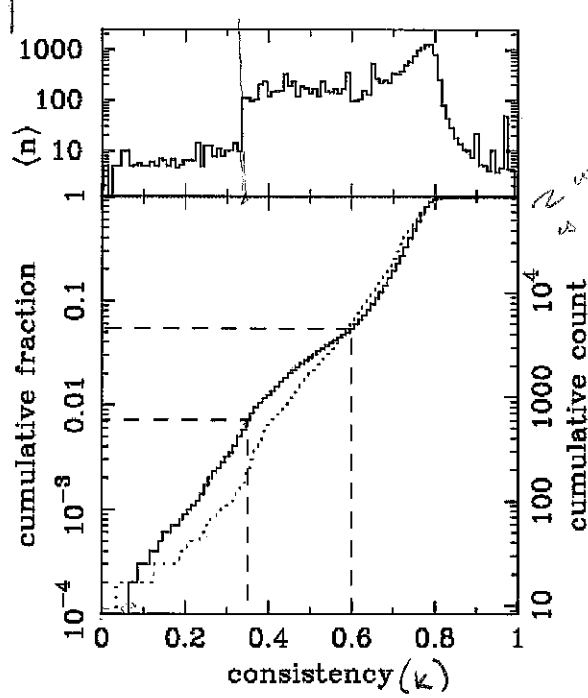

Figure 2. User consistency for Task XX. Top: total number of galaxies classified per user as a function of their consistency. Bottom: Cumulative fraction distribution of consistency. The dotted line shows the first iteration of weighting, and the solid line the third iteration. The second iteration is not shown, but is almost identical to the third.

what "fraction" is being accumulated? fr? Then should be consistency for Response X for Task Y.

what are the dashed straight lines

tive weighting scheme. First, we calculated the vote fraction ($f_r = n_r/n_{task}$) for every answer for every task for every object, weighting each user's vote equally. Here, n_r is the number of clicks for a given answer and n_{task} is the total number of clicks for that task. Individual clicks are then compared to the vote fraction to calculate its consistency κ:

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

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

$$\kappa_i = \begin{cases} f_r & \text{if click corresponds to this answer,} \\ (1-f_r) & \text{if click does not correspond.} \end{cases} \tag{2}$$

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

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

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

Based on the distribution of results for the initial iteration of κ (Figure 2), we chose a weighting function that down-weighted users in the tail of low consistency:

$$w = \text{power}((\kappa/0.6), 8.5) \tag{3}$$

For this function, $w = 1$ for $\sim 95\%$ of users and $w < 0.01$ for only $\sim 1\%$ of users. The vast majority of users are thus treated equally: there is no up-weighting of the most consistent users. The top panel of Figure 2 also shows the lowest-weighted users have on average classified only a handful of objects.

After computing κ for all tasks, the 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.

3.3 Classification bias

The weighted vote fractions in the data are also adjusted for what we term *classification bias*. The overall effect is a change in observed morphology fractions as a function of redshift, a trend seen in the original Galaxy Zoo 1 data. The presumed cause is that more distant galaxies are, on average, both smaller and dimmer as they appear in the cutout images; as a result, finer morphological features are more difficult to identify.

Figure 3 demonstrates the classification bias for several of the Galaxy Zoo 2 classification tasks. The average weighted vote fraction for each response (*thin lines*) 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) decrease 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 this effect is due to the nature of a luminosity-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 evidence of the classification bias even in magnitude-limited samples. Since this bias contaminates any potential studies of galaxy demographics over the entire volume of the sample, it must be corrected to the fullest possible extent.

Bamford et al. (2009) corrected for classification bias in the original Galaxy Zoo 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 measure the average elliptical-to-spiral ratio for each (M_r,R_{50}) bin in the lowest available redshift slice; 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 derive a correction for each (M_r,R_{50},z) bin and then adjust the vote fractions for the individual galaxies in each bin. The validity of this approach is justified in part by the agreement of these debiased probabilities with a monotonic morphology-density relation (Bamford et al. 2009). We modify and extend this technique for the Galaxy Zoo 2 classifications.

There are two major differences between the GZ1 and GZ2 data. First, GZ2 has a decision tree, rather than a single question and answer for each click on an image. This means that all tasks, with the exception of the first, depend on answers to previous classifications in the 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 weighted vote fraction for this example task only addresses the total bar fraction *among face-on disk galaxies*, and not as a function of the general population.

Our approach is to examine only biases within the context of the individual classification tasks. The corrections used to debias each task are derived based only on galaxies with sufficient votes to characterize that feature. We employ a combination of threshold on the weighted vote fraction for preceding tasks as well as a lower limit on the total number of votes for a galaxy to be used in deriving a correction. While this increases the number of noisy bins, it is critical for reproducing accurate baseline measurements of individual morphologies. The adjustment derived from well-classified galaxies is then applied to the vote fractions for *all* galaxies in the sample.

The second major issue is the adjustment of the GZ1 vote fractions assumed that the single task was essentially binary. Since almost every vote in GZ1 was either for "elliptical" or "spiral" (either anticlockwise or clockwise), they were able to use that ratio as the sole metric of the morphology. No systematic debiasing was done for the other GZ1 response options ("star/don't know", "merger", or "edge on/unclear"), and the method of adjusting the vote fractions assumes that these do not significantly affect the classification bias for the most popular responses.

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 morphologies for a given task. We represent this as the ratio of vote fractions (f_i/f_j) for responses i and j. Finally, 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:

really ?? is this a known astrophysical assumption? Isn't this just as bad as a proxy?

sample?

$$\frac{p_i}{p_j} = \frac{f_i}{f_j} \times K_{j,i} \tag{4}$$

confusion with consistency

In this case, the adjusted likelihood for a single task is written as:

$$p_i = \frac{1}{1/p_i}, \tag{5}$$

are these the same?

and the sum of all the likelihoods for the responses in a given task must be unity:

??

$$p_i + p_j + p_k + \cdots = 1. \tag{6}$$

Multiplying (5) by (6) yields:

$$p_i = \frac{1}{1/p_i} \times \frac{1}{p_i + p_j + p_k + \ldots} \tag{7}$$

$$p_i = \frac{1}{p_i/p_i + p_j/p_i + p_k/p_i + \ldots} \tag{8}$$

$$p_i = \frac{1}{\sum_{j \neq i} (p_j/p_i) + 1} \tag{9}$$

$$p_i = \frac{1}{\sum_{j \neq i} K_{j,i}(f_j/f_i) + 1}. \tag{10}$$

The corrections for each pair of tasks can be directly determined from the data. At the lowest sampled redshift bin ($z \simeq 0$), $\frac{p_i}{p_j} = \frac{f_i}{f_j}$ and $K_{j,i} = 1$. From Equation 4:

$$\left(\frac{f_i}{f_j}\right)_{z=0} = \left(\frac{f_i}{f_j}\right)_{z=z'} \times K_{j,i} \tag{11}$$

inverted

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

$$\tag{13}$$

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

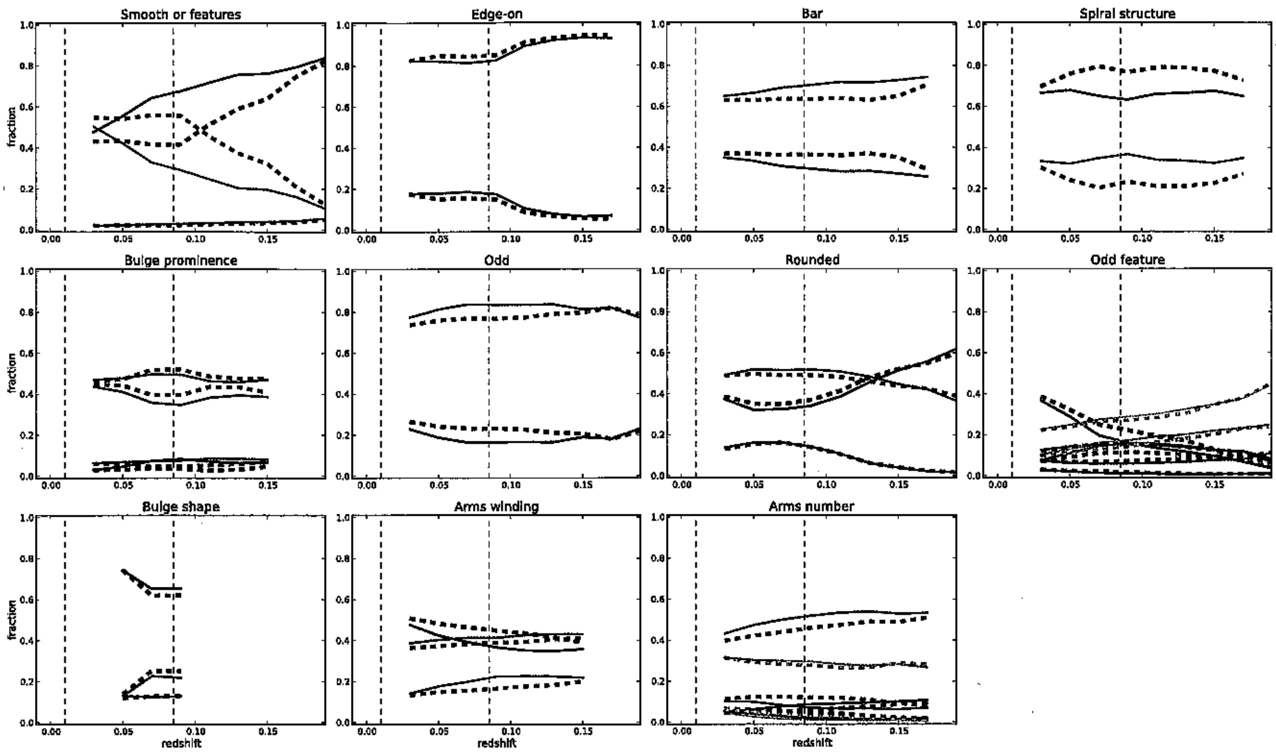

Figure 3. Type fractions for the classification tasks in GZ2. Solid (thin) lines show the weighted vote fractions, while the thick (dashed) lines show the debiased vote fractions that have been adjusted for classification bias. This is a magnitude-limited sample for $M_r < -20.17$. Vertical dashed lines 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).

inverted

$$C_{j,i} = \log\left[\left(\frac{f_i}{f_j}\right)_{z=z'} \Big/ \left(\frac{f_i}{f_j}\right)_{z=0}\right] \tag{14}$$

$$C_{j,i} = \log\left(\frac{f_i}{f_j}\right)_{z=z'} - \log\left(\frac{f_i}{f_j}\right)_{z=0}. \tag{15}$$

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

The local baselines and subsequent corrections are derived from the main sample data (original + extra + magnitude-limited Stripe 82). 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 only galaxies with sufficient numbers of classifications to determine the morphology ratios. This varies as a function of the task – for the questions asked of every galaxy (Tasks 01 and 06), we set the minimum number of classifications at 30. This is well below the median of 43, and includes > 97% of the sample. For other tasks with fewer total responses, this can be as low as 10 classifications per task.

The weighted vote fractions for each task response are binned in three dimensions: the absolute magnitude M_r, the Petrosian r-band half-light radius R_{50}, and redshift z. Bins ~~range~~ 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. The bin ranges and step sizes are chosen to maximize the phase space covered by the bias correction, while also retaining enough galaxies in each bin to establish its morphology distribution. The value of each bin in the cube is the sum of the weighted vote fractions for that response. For each pair of responses (i, j) to a question, we compute $\log(f_j/f_i)$ in every (M_r, R_{50}, z) bin. The local baseline relation is established by selecting the value in the non-empty bin(s) for the lowest-redshift slice at a given (M_r, R_{50}).

Since each unique pair of responses to a question will have a different local baseline, there are $\binom{n}{2}$ corrections for a task with n responses. This reduces to the method with a single pair of variables described in Bamford et al. (2009) if $n = 2$.

The baseline morphology ratios for the GZ2 tasks are shown in Figure 4 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. It is reasonably well-fit with an analytic function of the form:

$$\frac{f_j}{f_i}[R_{50}, M_R] = \frac{s_6}{1 + exp[(x_0 - M_R)/x_1]} + s_7 \tag{16}$$

where: