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 makes 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`.

Key words: galaxies

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` – now archived and not collecting data anymore.

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

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).

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 Img-Cutout 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

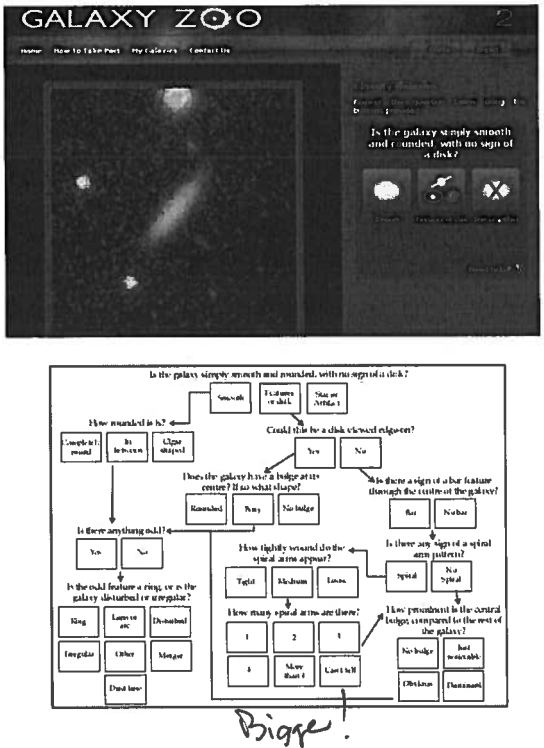

Figure 1. *Top:* Front page of the web interface for Galaxy Zoo 2, displaying Task 01. *Bottom:* Flowchart of the 11 classification tasks for GZ2, beginning at the top center.

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.

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 objects ($\sim 1\%$), and an even smaller proportion ($\sim 0.01\%$) 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 itera-

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 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 galaxies with sufficient votes to characterize that feature. We emply 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:

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

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

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

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

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

Multiplying (5) by (6) yields:

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

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

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

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

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:

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

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

$$\tag{13}$$

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

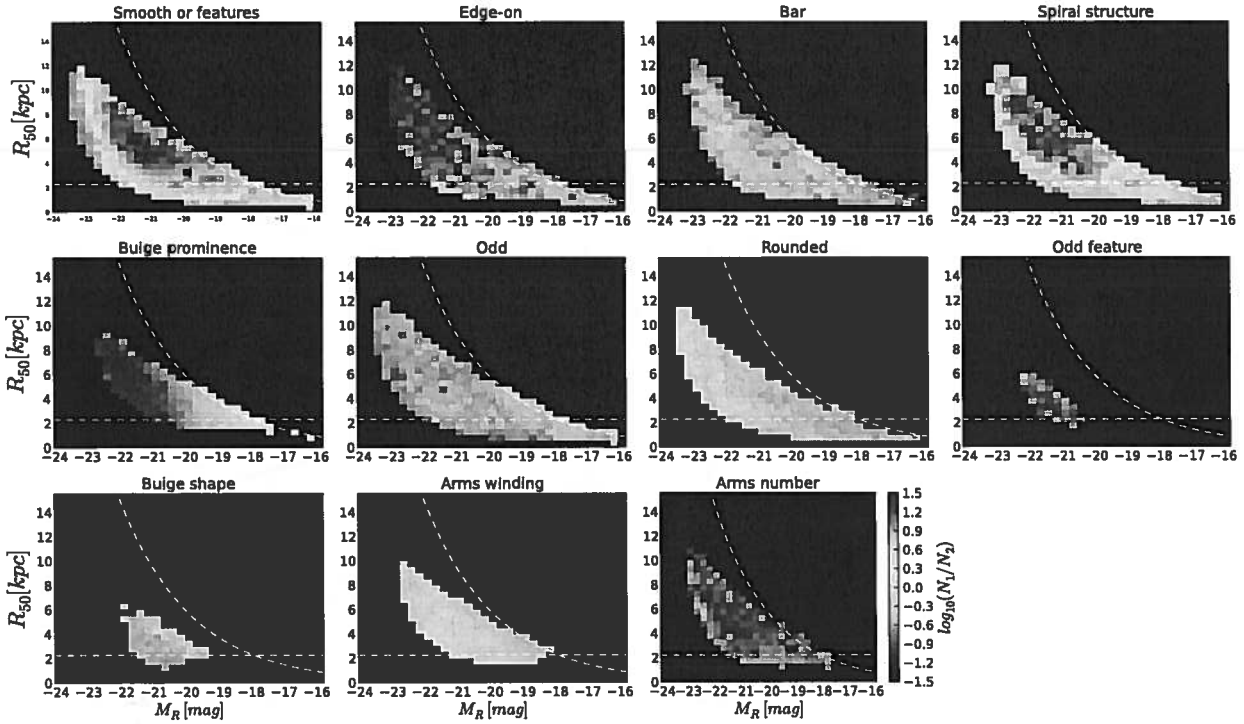

Figure 4. Local morphology ratios for morphology classifications in GZ2. The ratio of the binned vote fractions for morphologies is for the first two responses in the decision tree (Table 2) for each task; there may be as many as 21 such pairs for tasks with more than two options. The dashed horizontal 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}.

$$\alpha = s_2^{-(s_1+s_8 R_{50}{}^{s_9})} + s_3 \quad (17)$$

$$\beta = s_4 + s_5(x_0 - s_3) \quad (18)$$

and where $\{s_1, s_2, s_3, s_4, s_5, s_6, s_7, s_8, s_9\}$ are minimized to fit the data. The only other task had baseline ratios reasonably well fit by these parameters was Task 07 (rounded 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 16; for these, we instead adopt a simpler fit where both M_r and R_{50} vary linearly:

$$\frac{f_j}{f_i}[R_{50}, M_R] = t_1(R_{50} - t_2) + t_3(M_R - t_4) + t_5, \quad (19)$$

and $\{t_1, t_2, t_3, t_4, t_5\}$ are the parameters to be minimized. We fit Equation 19 to all other tasks where the number of bins is sufficient to get a reasonable fit. Finally, for pairs of responses with only a few sampled bins, we instead used the direct 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 generally good for most GZ2 tasks and responses. Figure 3 illustrates the comparison between the raw and debiased vote fractions. The debiased results (*thick lines*) are generally flat over a range of $0.01 < z < 0.085$, where L^* galaxies fall below the magnitude limit of the survey and the bins are more poorly sampled. 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). The bar fraction in disk galaxies is roughly 0.35, which is slightly higher than the value found by using thresholded GZ2 data in Masters et al. (2011).

4 THE CATALOG

Other possible inclusions for catalog:

- Metrics on classification confidence (Table 04, Lintott et al. 2011)
- Galaxy Wars
- M_r,R_{50},z bins for each galaxy
- Voronoi tessellation bins
- Matched SDSS metadata ← don't you just need the objid?

4.1 Main sample

The data release for Galaxy Zoo 2 consists of four tables, abridged portions of which appear in this paper. Table 3 contains classification data for the 283,971 galaxies in the main sample. Each galaxy is identified by its unique SDSS DR7 objID, as well as its original sample

fraction for each response. The distributions for both the main sample and Stripe 82 galaxies are quite similar, with the difference in the mean varying by $< 10\%$ for almost all responses. The only exceptions to these are for responses that target rare objects (and thus are subject to higher variance for low-number statistics), such as dust lanes, rings, and high-multiplicity spiral arms.

The right panel of Figure 5 shows that the weighted vote fractions also behave similarly as a function of redshift, particularly in the $0.01 < z < 0.08$ range covered by the GZ1 debiasing technique. The agreement is generally good between the Stripe 82 normal depth and the GZ2 main sample; this is not the case for the coadded Stripe 82 data, however. For Task 01, fewer galaxies are classified as robustly smooth (above the 0.8 threshold), moving instead to the "unclassified" category. Coadded data showed similar higher fractions of galaxies with bars and for possessing visible spiral structure. A possible cause for this is that the new image pipeline in the coadded data allows viewers to see faint features or disks, due to improved seeing in the coadded data (from $1.4''$ to $1.1''$; Annis et al. 2011).

For almost every response in the GZ2 decision tree, the data (no bias correction) have no systematic differences between classifications using the coadd1 and coadd2 images. Figure 6 shows distributions of the differences between the two weighted vote fractions ($\Delta_{coadd} = f_{coadd1} - f_{coadd2}$). If the mean value of Δ_{coadd} for an answer is 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 Task 05, Answer 11 (prominence of the bulge is "just noticeable"), for which the mean weighted fraction in coadd2 data is $\sim 35\%$ higher than from coadd1 data. This is an opposite (but not equal) effect than Answer 12 (obvious bulge), for which the coadd1 data is $\sim 13\%$ 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 unknown.

The comparison of the coadd1 and coadd2 data sets also demonstrates the intrinsic variability in classification of a single object, even with several tens of votes. For example, in the (unbiased) vote fractions from Task 01, 6831 (32.0%) galaxies from coadd1 and 7,244 (33.9%) galaxies from coadd2 exceed the "clean" early-type threshold of $p \geqslant 0.8$. However, only 2,300 galaxies meet this threshold in *both* samples, while the union of the two yields 11,602 galaxies. The difference in numbers between the samples decreases when a higher value of p is used; a more robust jackknife sampling of the data would improve on this 1-sample jackknife.

Type fractions for both weighted and debiased vote fractions in normal-depth Stripe 82 are shown in Figure 7 for a subset of the GZ2 tasks. The correction flattens the redshift effect in all tasks, similar to the main sample data. The variance along redshift bins is somewhat higher – formal error bars will need to be computed to see if there is any statistical difference, or whether the result is consistent with a smaller total sample of galaxies.

4.3 Additional data

Although not reproduced in this paper, the repository at `http://data.galaxyzoo.org` contains pre-matched tables containing SDSS metadata for the spectroscopic galaxies in the GZ2 main sample. This contains some of the most commonly used DR7 parameters including SDSS exposure information, position, photometry, size, and redshift. Rows are matched to the corresponding galaxies in Tables 3–5. These are provided as a resource for members of the community who wish to compare the morphogical data against external parameters.

5 COMPARISON OF GZ2 TO OTHER CLASSIFICATION METHODS

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

5.1 Galaxy Zoo 1 vs. Galaxy Zoo 2

As a check of the classification accuracy, we compare the results from GZ2 to those in GZ1 (Lintott et al. 2011). The galaxies in GZ2 are a subset of those in GZ1, with 248,883 matches between the samples. Task 01 in GZ2 is almost identical to the interface of GZ1. GZ1 allowed for selection of "merger" and "don't know" options in addition to the first three; and asks for galaxies with "features or disk" rather than only for spiral structure.

The matched GZ1-GZ2 catalog contains 34,480 galaxies flagged as "clean" ellipticals based on their debiased GZ1 likelihoods. Of those, 89.0% had GZ2 raw vote fractions greater than 0.8 and 99.9% greater than 0.5. Using the GZ2 debiased likelihoods, the vote fractions match at 50.4% at a threshold of 0.8 and 97.6% at a threshold of 0.5.

There are 83,956 galaxies identified as "clean" spirals in GZ1. The agreement with the "features or disk" response in GZ2, however, is significantly lower. Only 31.6% of the GZ1 clean spirals had GZ2 raw 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 catalogs. The weighted vote fractions show a tight correlation at both very low and very high values of f_{sp}, indicating that both projects agree on the strongest spirals (and corresponding ellipticals). At 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 f_{sp}.

Based on the vote fractions, GZ2 is significantly more conservative than GZ1 at identifying spiral structure. One

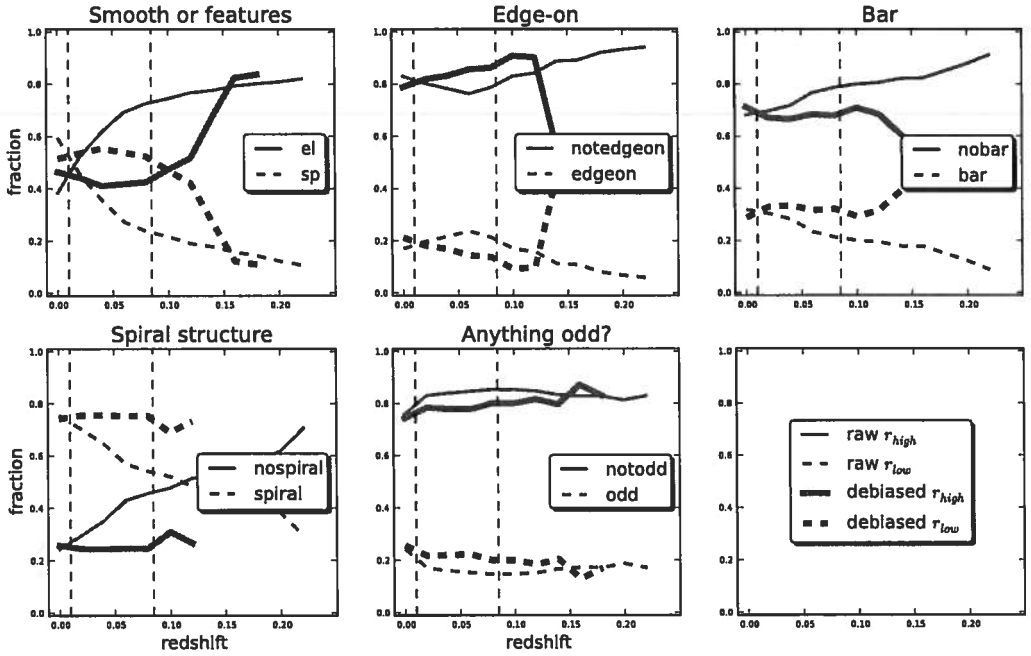

Figure 7. Type fractions for five binary GZ2 tasks for the normal-depth Stripe 82 data. This is a magnitude-limited sample for $M_r < -20.17$. Vertical dashed lines show the redshift at $z = 0.01$ (the lower limit of the correction) and $z = 0.085$ (the redshift at which the absolute magnitude limit reaches the sensitivity of the survey).

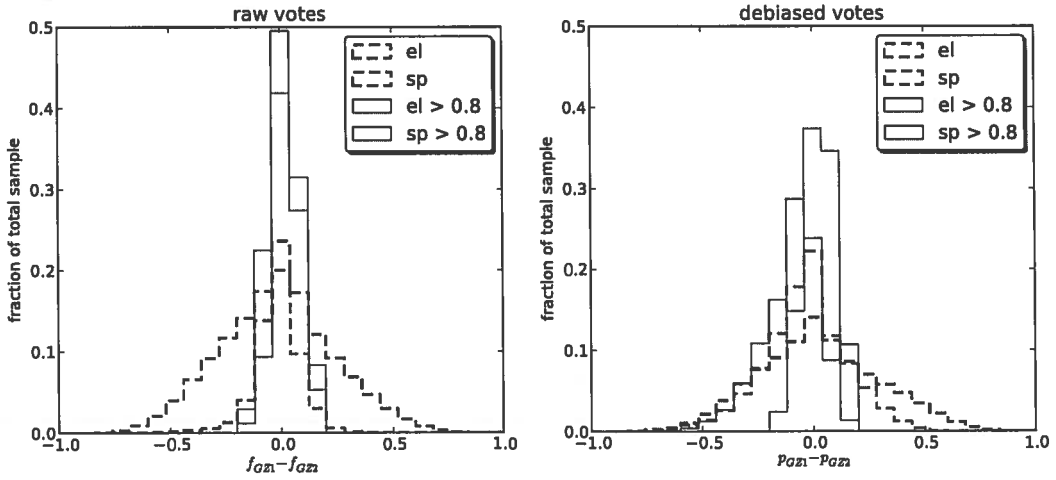

Figure 9. Differences in the vote fractions for galaxies in both the Galaxy Zoo 1 (GZ1) and Galaxy Zoo 2 (GZ2) projects. Left: Distribution of the differences in the raw weighted vote fractions. Dashed lines show data for all galaxies, while solid lines are for the subset in which f_{el} or $f_{sp} > 0.8$ in both samples. Right: same plot, but using the debiased vote fractions for both samples.

smooth and "features or disk" data for GZ2. For the raw vote fractions, galaxies showed a significant skew toward being 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

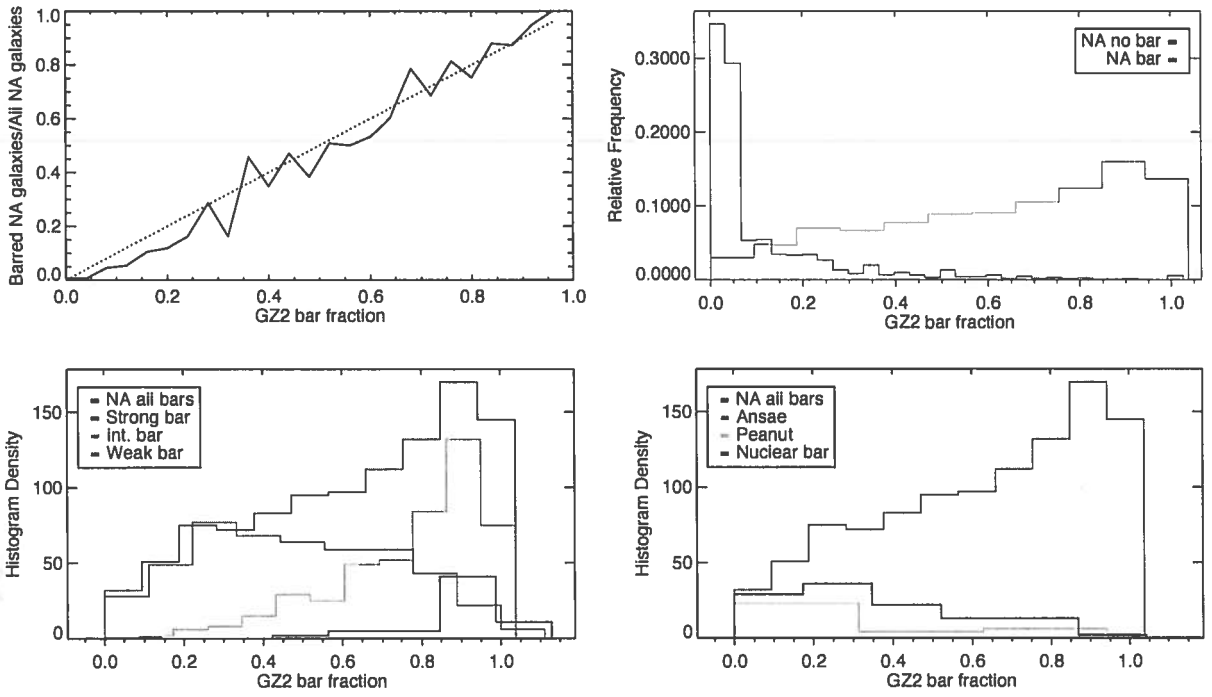

Figure 10. Galactic bar classifications for GZ2 and NA10. Data are for the 7,121 galaxies which are face-on (log $(a/b) < 0.3$), have 10 or more GZ2 bar classifications, and appear in both samples. Top left: mean weighted bar fraction per galaxy in GZ2 vs. the ratio of barred to all galaxies in NA10 (*solid line*). Dashed line shows the one-to-one relationship. Top right: distribution of the GZ2 bar vote fraction, separated by NA10 classifications (including all bars). 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 ansae, peanut, and nuclear bars in NA10.

The lack of sensitivity to medium and weak bars (as IDed by NA10) may be due in part to the design of the GZ2 interface. When users are asked if a bar is present, they are presented with a simple pictogram showing two examples of a barred galaxy. For the pictogram showing a disk, the bar drawing extends across its full diameter, fitting with the typical definition of a strong bar. With this as the only example (and no continuum of options between the two choices), GZ2 users may not have looked for bars shorter than the disk diameter, or been less confident in clicking "yes" if they did see them. Results from Hoyle et al. (2011) show that users are fully capable of identifying weak bars in other contexts; however, the construction of our decision tree means that GZ2 classifications only include examples from strong and medium bars.

NA10 identify three other fine-structure features related to bars: ansae, peanuts, and nuclear bars. None of the three correlate strongly with the GZ2 bar parameter, with more galaxies actually having vote fractions < 0.5 than above it. 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. No ansae are detected in the GZ2 face-on subsample, likely due to our axial cut.

In the GZ2 main sample, 15,873 galaxies are in the clean, debiased sample of bars (as set by the flag in the data). This is 29.6% of the clean sample of face-on galaxies, in extremely good agreement with both the 29.4% fraction found in the smaller volume-limited sample of Masters et al. (2011), and with the $\sim 30\%$ for disk galaxies with $(b/a) > 0.4$ (or log $(a/b) < 0.398$) and T-types of S0 or later of Nair & Abraham (2010b).

5.2.2 Rings

NA10 also classifies ring galaxies in their catalog. They include three basic types of rings based on the Buta & Combes (1996) criteria. Inner rings lie between the bulge and spiral arms or disk. Outer rings are external to the spiral arms, but are still closely linked to the spiral pattern. Nuclear rings lie in the bulge region of galaxies; no specific size scale for this is given. In GZ2, rings are classified only if the user answers "yes" to the question *"Anything odd?"* Since the user then has seven different options to choose from (ring, lens, disturbed, irregular, other, merger, dust lane), interpreting the weighted ring fractions is less straightforward.

In the NA10 catalog, 18.2% of galaxies have a ring. Of those, 10% are nuclear rings, 74% are inner rings, and 32% are outer rings (sum is more than 100% since $\sim 1/3$ of ringed galaxies have multiple rings flagged). In the GZ2 catalog, 3,142 galaxies are in the clean sample of rings, but this is based on only a potentially small number of total votes ($N \geqslant 5$). In both catalogs, selecting only face-on galaxies did not significantly change the percentage of galaxies identified as having a ring.

In the top-left of Figure 11, the distribution of the num-

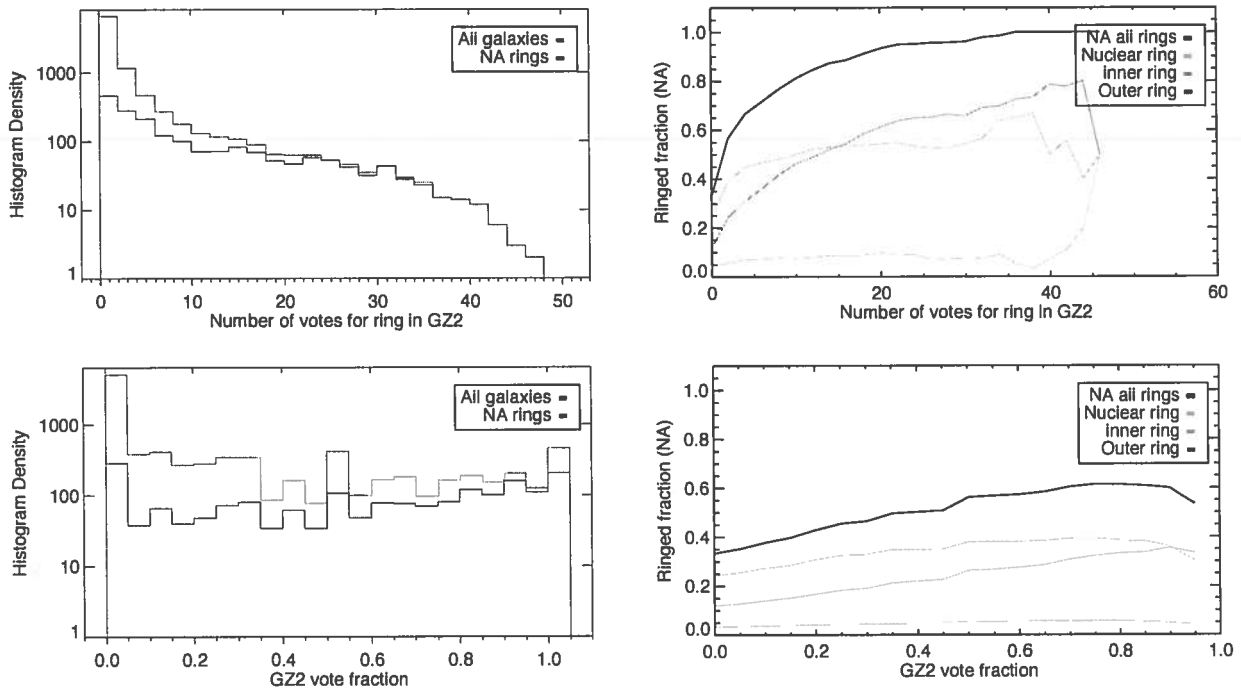

Figure 11. Ring classifications in GZ2 and NA10. Data are for the 9,746 galaxies in both samples which are face-on (log $(a/b) < 0.3$). Top left: distribution of NA10 ringed galaxies compared to all face-on galaxies as a function of the raw number of votes for a ring in GZ2. Bottom right: distribution of NA10 ringed galaxies compared to all face-on galaxies as a function of the GZ2 vote fractions for ring. Top right: Fraction of face-on galaxies that have a ring (NA10) as a function of the total number of votes for a ring in GZ2. Bottom right: Fraction of face-on galaxies that have a ring (NA10) as a function of the GZ2 ring vote fraction.

0.8, for example, would include 4% late-type galaxies (20% if S0 galaxies are included).

For galaxies identified as having features that are not edge-on disks, 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, $\sim$ 85% of them have GZ2 weighted fractions of 0.0, with the remainder weakly clustered around 0.3. For NA10 late-type galaxies, the majority of disk/feature objects have high GZ2 spiral structure weighted fractions. For galaxies with at least 10 votes on Task 04 (a peak at 0.0 appears when this cut is not imposed), 70% of Sa or later-types have a GZ2 spiral vote fraction > 0.8. This drops to 60% if S0 galaxies are included as late-type. The missing population is thus made up of galaxies with significant spiral structure by NA10, 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. Experimenting with other values for the GZ2 weighted fraction had no change on the results.

5.2.5 *Spiral tightness*

If a disk galaxy was identified as having 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 pictograms on buttons that illustrated representative pitch angles). This has direct (but not exclusive) connections to the Hubble classification of late-type galaxies; tight spirals would 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 14 shows the distribution of NA10 T-types for galaxies based on their GZ2 weighted fractions for winding arms. This figure shows only galaxies with at least 10 votes on spiral structure; looking at all galaxies in the overlapping sample disproportionally weights the 0.0 and 1.0 weighted fraction bins. Weighted 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 weighted fraction of less than 0.2. Almost no elliptical galaxies from the NA10 catalog are included, although there are significant numbers of S0 galaxies.

For tight spirals, galaxies with the highest weighted fractions have more earlier-type spirals than galaxies with a low vote for tight spiral winding arms. For a tight spiral weighted fraction above 0.9, 85% of galaxies are Sb or

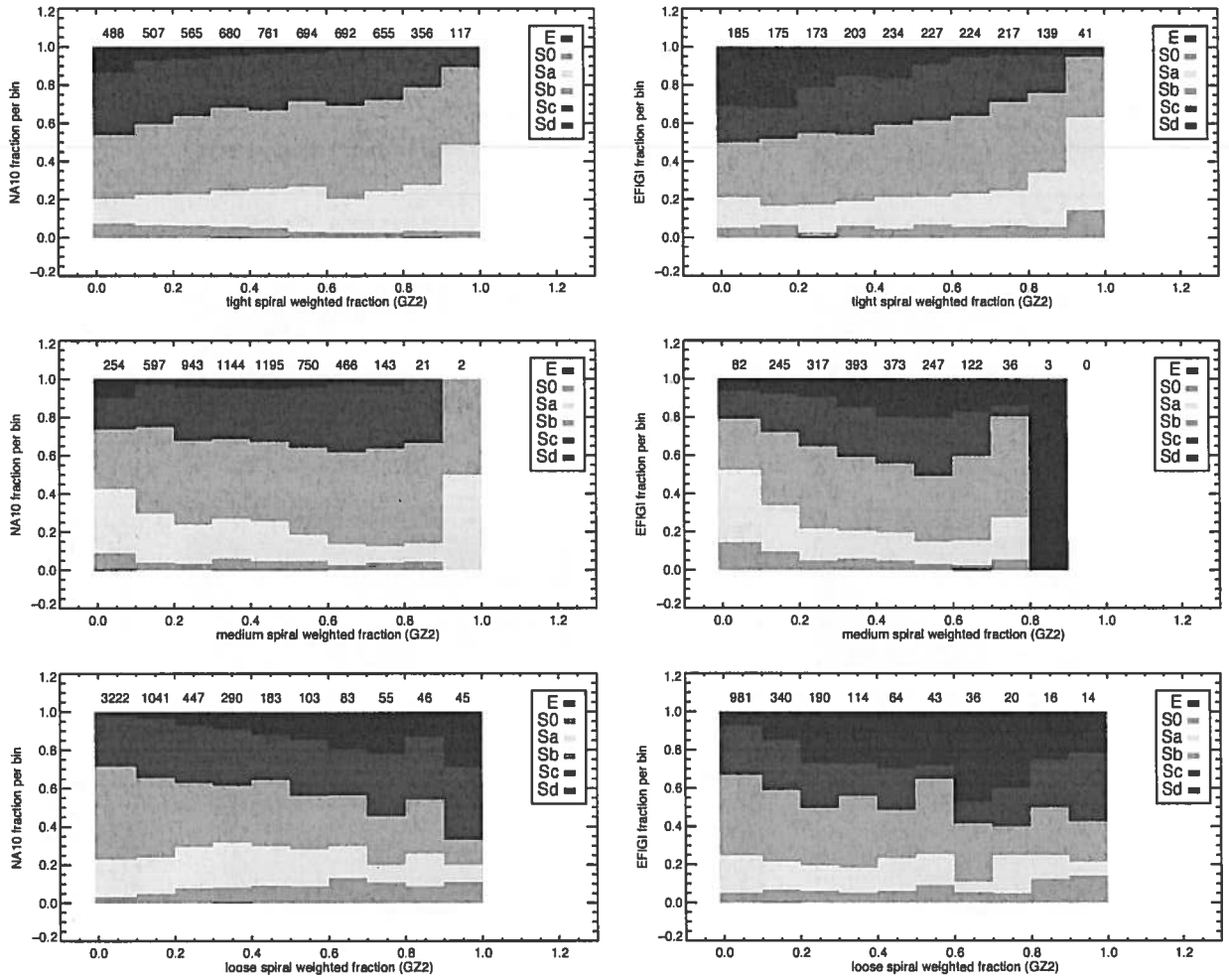

Figure 14. T-type classifications compared to the GZ2 vote fractions for spiral tightness (Task 10). Left side is NA10 T-types; right side is EFIGI T-types. Data are for the 5,515 (NA10) and 1,907 (EFIGI) galaxies, respectively, with at least 10 GZ2 votes for Task 10. The number of galaxies per bin is given along the top of each panel.

earlier. Medium-wound spirals with high weighted fractions tend to be Sb and Sc galaxies – the proportion of both types increases as a function of medium-wound weighted fraction, and constitute 84% of galaxies when the weighted fraction is greater than 0.6. Galaxies strongly classified as medium-wound are rare, however, with only 23 galaxies having a weighted fraction above 0.8. Loose spirals are dominated by Sc and Sd galaxies at high weighted fraction values, comprising more than 50% of galaxies above a loose weighted fraction of 0.7.

Overall, we see a clear trend for looser GZ2 spiral arms to correspond with later spiral T-types from NA10 classifications. High weighted vote fractions are mostly Sa/Sb galaxies for tight winding, Sb/Sc galaxies for medium winding, and Sc/Sd galaxies for loose winding. Individual GZ2 vote fractions, however, have significant diversity even at the highest bins, and do not reliably separate the morphologies on the level of the Hubble T-types.

5.2.6 Bulge dominance

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). Similar to the arm tightness task, we analyze the relationships between GZ2 classification and Hubble types from NA10

The left side of Figure 15 shows the distribution of NA10 T-types for galaxies based on their GZ2 weighted fractions for winding arms. This figure shows only galaxies with at least 10 votes on bulge prominence. Weighted 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

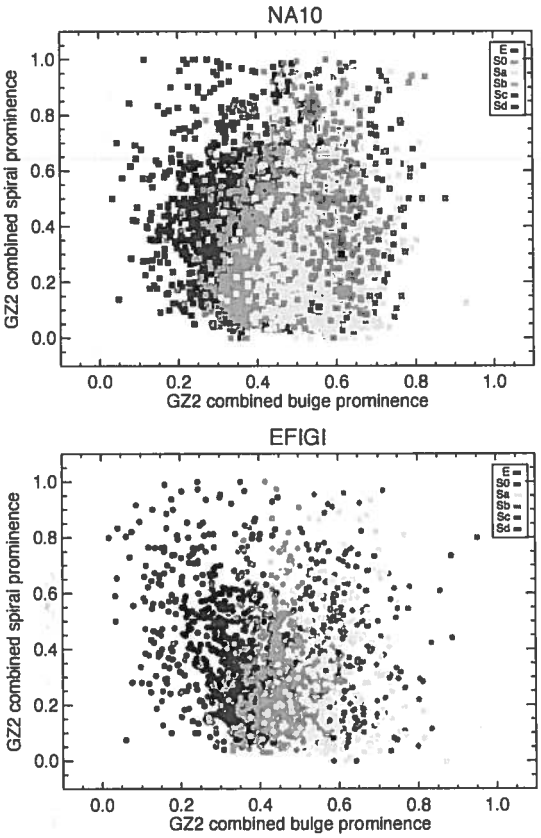

Figure 16. Weighted bulge classifications vs. weighted spiral classifications from GZ2. Galaxies are color-coded by their morphologies in EFIGI (*top*) and NA10 (*bottom*). Data are for galaxies with at least 10 classifications for both Task 05 (bulge dominance) and Task 10 (spiral structure) in GZ2.

$$\omega_{spiral} = (0 \times f_{tight} + 1 \times f_{medium} + 2 \times f_{loose})/2. \quad (21)$$

Using the weighted feature classifications shows a clear separation in T-types; the majority of this, however, is for the bulge prominence category. Spiral prominence does not seem to significantly affect any of the classifications.

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.

5.3 EFIGI

Baillard et al. (2011) performed individual morphological classifications of 4,458 galaxies for EFIGI (Extractions de Formes Idéalisées de Galaxies en Imagerie). The sample is a subset of the RC3 catalog for which 5-color imaging in the SDSS DR4 was available. Images are supplemented by redshift information from several different sources. The galaxies have no strong redshift or volume limit on the sample, with almost all galaxies at $0.0001 < z < 0.08$. Classifications on composite *gri* images were performed by a group of 11 salaried astronomers, each of whom classified a subset of 445 galaxies. A training set of 100 galaxies was also classified by all astronomers in the group to adjust the sample for individual bias.

EFIGI contains two types of morphological classification: T-types and attributes. T-types are assigned using a slightly modified version of the RC3 Hubble classifications. Peculiar galaxies are not considered a separate stage, and ellipticals are subdivided into various types: compact, elongated (standard elliptical), cD (giant elliptical), and dwarf spheroidals. They also classify late-type lenticulars ($S0^+$; T-type=-1) that are not included in the classification of NA10. The remaining morphological information, called 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`

Attributes are defined on a five-step scale from 0 to 1 (0, 0.25, 0.50, 0.75, 1) that describe the strength of the feature in question. 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; this scale may not be linear. For others (eg, inclination or multiplicity), it ranges between the extrema of possible values. A 70% confidence interval (roughly 1σ) is estimated by setting lower and upper limits on the same five-point scale.

EFIGI is compared in detail to NA10 in Baillard et al. (2011). Only $\sim$ 10% of the NA10 catalog overlaps with EFIGI classifications; roughly one-third of the EFIGI sample lies at redshifts below the NA10 lower limit of $z = 0.01$, and also contain significant number of galaxies fainter than $g = 16$. T-types agree well between the two samples; 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 that have no interaction flags in the NA10 catalog, 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.

3,411 galaxies appear in both EFIGI and GZ2. This constitutes 77% of the EFIGI galaxies and 1.2% of the GZ2 sample. Like NA10, it offers a supervised set for comparisons between citizen scientists and salaried astronomers. The comparisons also benefit (or possibly detract) from the fact that EFIGI, like GZ2, has multiple classifiers and thus possible variance based on individual bias.

5.3.1 Bars

GZ2 asks users to identify whether a bar is present in the galaxy. EFIGI's scale is based on bar length (not necessarily

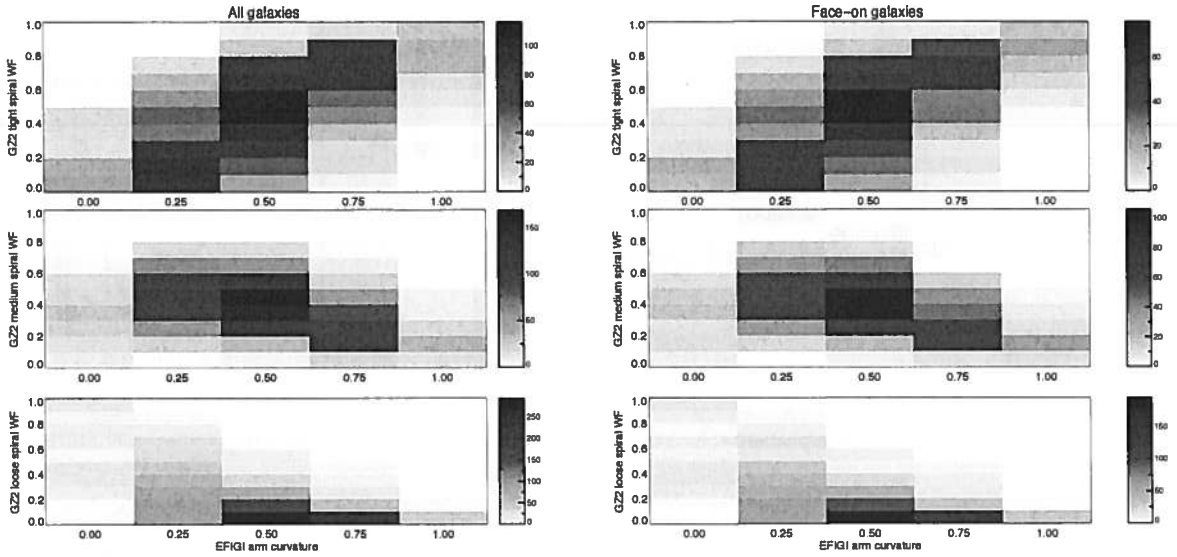

Figure 18. EFIGI arm curvature classifications compared to their GZ2 weighted fractions for the presence of a bar. Data on the left are for the 3,411 galaxies in both samples; the subset of 2,099 face-on galaxies is on the right. Dashed lines on the top and bottom pairs of plots show the one-to-one correlation, a Gaussian with $\mu = 0.5$ and $\sigma = 0.25$, and the one-to-one anti-correlation, respectively.

patterns for all three categories. The correlation coefficient for both tight and medium spirals does decrease to $|\rho < 0.2|$ if a 10-vote lower limit is applied.

Notes:

- Most useful: look at high weighted fractions and high numbers of vote counts for the spiral winding category of choice. That might correlate most strongly with EFIGI values.
- EFIGI values are matched to specific pitch angles of the galaxy. For galaxies in which the GZ2 WF does not match the EFIGI, what is happening? Are the vote categories laterally skewed, are we not sensitive to weakest spirals, or is something else going on?
- Not convinced that the face-on criterion is working.
- Change bar plot to single-panel version. Change right side of arm curvature to cuts on count and weighted vote fraction, show stronger correlations. Discuss the percentage of the total number of galaxies that this constitutes for the expanded GZ2 sample, and how many "clean" galaxies from the EFIGI criteria we might be able to extract (subject to bias at lower surface brightnesses).

5.4 Huertas-Company – automated classifications

Huertas-Company et al. (2011, HC) published a study in which they used training sets of galaxy images to create automated classifications, and then compared the results to GZ1. There have been no comparisons to GZ2; however, the broad nature of their probabilities (four broad morphological categories) largely unsuited for the fine structure questions such as bar and spiral number.

The sample of galaxies classified by HC is the SDSS DR7 spectroscopic sample, limited to galaxies with $z < 0.25$ with good photometric data and clean spectra. Their total of 698,420 galaxies is approximately twice the size of the GZ2 sample, which has a similar redshift range. The HC sample goes to fainter magnitudes, with more than 400,000 galaxies below the GZ2 limit of $r > 17$. Their algorithm is implemented using support vector machine software that tries to find boundaries between points in N-dimensional space, where N is determined by criteria including morphology, luminosity, color, and redshift (Huertas-Company et al. 2008). The training set is the 2,253 galaxies in Fukugita et al. (2007), which have 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 combine their two respective late-type categories).

Huertas-Company et al. (2011) directly compare their results to the GZ1 sample from Lintott et al. (2011). They find 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 catalog. They also find a (but not strictly linear) relationship between the GZ1 debiased vote fraction and the HC probabilities. This is one of the first independent confirmations that the vote fractions may be related to the actual *probability* of a galaxy possessing a particular morphology.

HC also compare their results to the NA10 data, most of which are not included in the HC training sample. They find a good correlation between the NA10 T-types and the HC probabilities, especially for ellipticals and Scd spirals. S0 galaxies are more difficult to separate; for the NA10 lenticulars, HC11 give only a 0.4 probability of S0, with 0.32 of elliptical and 0.2 of being Sab. Sab galaxies have an average probability of 0.55 being an Sab in HC, but also 0.15

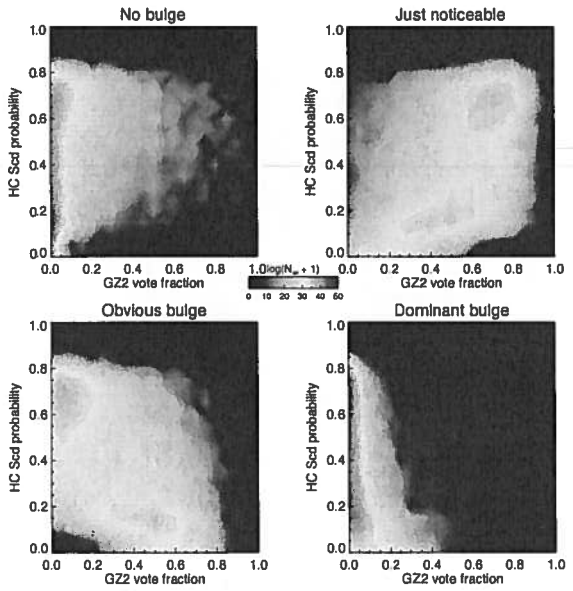

Figure 22. Huertas-Company et al. (2011) early-type probability as a function of the GZ2 weighted vote fraction for bulge dominance. The color of the contours is log $(N_{gal} + 1)$, where N_{gal} ranges from 0 to 1.5×10^3. Only galaxies with at least 10 classifications for Task 05 are shown.

- Check how the presence of a bar affects the HC classifications.
- Analyze the HC T-type vs. GZ2 bulge prominence (Figure 22).

5.5 Other classification schemes

Mention the volume classification of de Vaucouleurs (1959); recent developments by Kormendy & Bender (2012); Laurikainen et al. (2011); Cappellari et al. (2011); Krajnović et al. (2011) that have refined these ideas. Buta (2011) is a useful review of modern morphology.

6 ASTRONOMY

Statistical properties and demographics of galaxies to be placed here. At minimum, a table of the number of galaxies in each clean sample would be useful, plus a color-magnitude diagram. Compare to early science results of EFIGI (de Lapparent et al. 2011) and NA10 (Nair & Abraham 2010b; Nair et al. 2010)

7 CONCLUSIONS

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 have been possible. Their efforts are individually acknowledged at `http://authors.galaxyzoo.org`.

Boilerplate for individual grants, Zooniverse support, etc.

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., Adelman-McCarthy J. K., Agüeros M. A., Allam S. S., Allende Prieto C., An D., Anderson K. S. J., Anderson S. F., Annis J., Bahcall N. A., et al. 2009, ApJS, 182, 543

Aguerri J. A. L., Méndez-Abreu J., Corsini E. M., 2009, A&A, 495, 491

Annis J., Soares-Santos M., Strauss M. A., Becker A. C., Dodelson S., Fan X., Gunn J. E., Hao J., Ivezic Z., Jester S., Jiang L., Johnston D. E., Kubo J. M., Lampeitl H., Lin H., Lupton R. H., Miknaitis G., Seo H.-J., Simet M., Yanny B., 2011, ArXiv e-prints

Baillard A., Bertin E., de Lapparent V., Fouqué P., Arnouts S., Mellier Y., Pelló R., Leborgne J.-F., Prugniel P., Makarov D., Makarova L., McCracken H. J., Bijaoui A., Tasca L., 2011, A&A, 532, A74

Bamford S. P., Nichol R. C., Baldry I. K., Land K., Lintott C. J., Schawinski K., Slosar A., Szalay A. S., Thomas D., Torki M., Andreescu D., Edmondson E. M., Miller C. J., Murray P., Raddick M. J., Vandenberg J., 2009, MNRAS, 393, 1324

Banerji M., Lahav O., Lintott C. J., Abdalla F. B., Schawinski K., Bamford S. P., Andreescu D., Murray P., Raddick M. J., Slosar A., Szalay A., Thomas D., Vandenberg J., 2010, MNRAS, 406, 342

Barazza F. D., Jogee S., Marinova I., 2008, ApJ, 675, 1194

Buta R., Combes F., 1996, FCP, 17, 95

Buta R. J., 2011, ArXiv e-prints

Cappellari M., Emsellem E., Krajnović D., McDermid R. M., Serra P., Alatalo K., Blitz L., Bois M., Bournaud F., et al. 2011, MNRAS, 416, 1680

Darg D. W., Kaviraj S., Lintott C. J., Schawinski K., Sarzi M., Bamford S., Silk J., Proctor R., Andreescu D., Murray P., Nichol R. C., Raddick M. J., Slosar A., Szalay A. S., Thomas D., Vandenberg J., 2010, MNRAS, 401, 1043

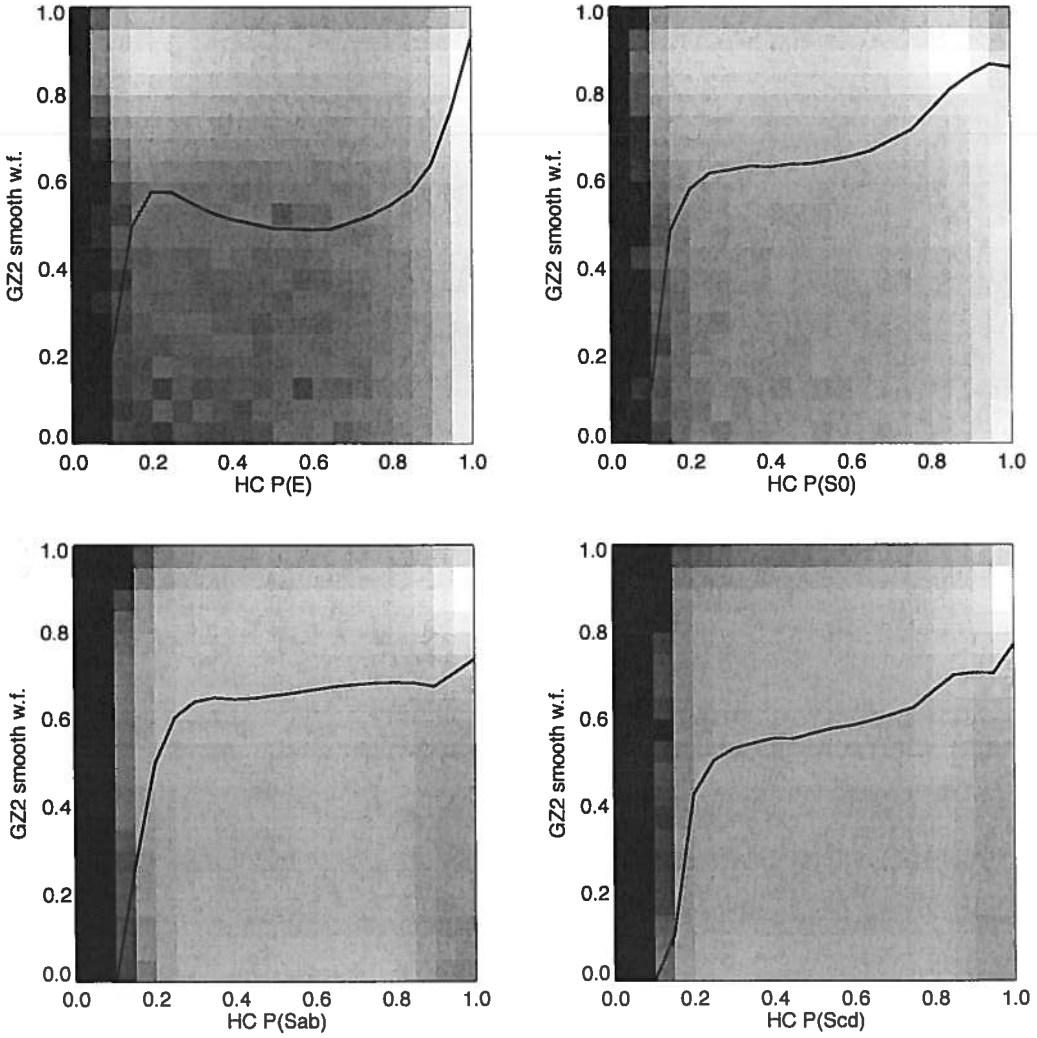

Figure 21. Left: GZ2 smooth weighted fraction as a function of Huertas-Company et al. (2011) early-type probability. Right: GZ2 features/disk weighted fraction as a function of HC late-type probability. Whiter values in both indicate a larger fraction of galaxies in that bin (logarithmic scale).

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

Table 5. Morphological classifications of GZ2 Stripe 82 coadded galaxies with spectra

			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	0.218	0	5	5.0	0.312	0.312	0.775	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	0.783	0	4	4.0	0.174	0.174	0.174	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	0.280	0	4	4.0	0.174	0.180	0.826	1	
8647474690312898532	26	129	12	12.0	0.462	0.462	0.462	0	14	14.0	0.538	0.538	0.538	0	
8647474690312962734	20	69	18	17.0	0.900	0.895	0.895	1	2	2.0	0.100	0.105	0.105	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 3. Classifications here are for the coadded images (version 2) from Stripe 82, which goes to a deeper magnitude limit and has a better angular resolution than galaxies in the main sample.